



Intelligent State Evaluation and Fault Diagnosis Methods for Wind Turbines

Guoliang Wang

Submitted for the degree of Doctor of Philosophy

Department of Electrical and Electronic Engineering

University of Sheffield

February 2025

Declaration

I solemnly declare that this dissertation titled " Intelligent State Evaluation and Fault Diagnosis Methods for Wind Turbines " is my work, completed under the meticulous guidance of Dr. Hua-Liang Wei at the University of Sheffield. I affirm that all sources utilized in this study have been duly acknowledged and cited. Furthermore, this dissertation contains no material that has been submitted for any other academic award or degree. I acknowledge that any assistance received during the research and writing of this dissertation has been appropriately acknowledged.

Guoliang Wang

February 2025

Acknowledgement

I extend my sincere gratitude to Dr. Hua-Liang Wei, my esteemed supervisor, for his invaluable guidance, unwavering support, and profound insights throughout the journey of this doctoral research. His expertise, encouragement, and dedication have been instrumental in shaping the trajectory of this study.

I am deeply thankful to the University of Sheffield for providing the conducive academic environment and resources for the completion of this research endeavor.

I would also like to express my appreciation to my colleagues and friends for their encouragement and assistance throughout this endeavor.

Lastly, I am profoundly grateful to my family for their unwavering love, encouragement, and understanding during this academic pursuit.

Abstract

Wind power is an important green and sustainable source of power generation. However, the construction of wind farms needs a large amount of initial investment and a highly expensive maintenance cost for wind turbines (WTs). Therefore, it is crucially important to accurately assess the state of WTs and keep WTs in good operation conditions. This thesis proposes a framework for the novel intelligent state evaluation and maintenance arrangement (iSEMA) system based on digital twin (DT) and deep learning (DL) technology, which can accurately evaluate the state of WTs and detect faults in the early stage. With the proposed iSEMA system, operators can better monitor WT conditions and use the information provided by the iSEMA system to schedule maintenance work more effectively, thereby minimizing downtime and boosting the economic efficiency of WTs.

In addition, this thesis creatively gives the concept of the sub-healthy state of WTs by introducing some relevant quantities, which is very useful for designing the iSEMA system and better describing the state of WTs. Based on the historical data of WTs, the iSEMA system can automatically select the suitable value of the parameters.

In the research about bearing fault diagnosis methods, this thesis proposed a novel hybrid framework by combining traditional machine learning (ML) and advanced DL models. The proposed hybrid framework can not only realize the automatic feature extraction for random forest (RF), but also significantly improve the accuracy for handling multi-classification problems.

Based on the auxiliary classifier generative adversarial networks (ACGANs) which is an advanced variant of the generative adversarial network (GAN), this thesis proposed an improved bearing fault diagnosis method for handling the imbalanced data problem. This method addresses the issue of insufficient fault data in real-world scenarios, which leads to low accuracy of data-driven methods in fault diagnosis.

Table of Contents

Abstract	III
LIST OF FIGURES	VII
LIST OF TABLES	IX
Abbreviations	XI
Introduction	1
1.1 Background and Motivations	1
1.2 Aims and Objectives	3
1.3 Research Gaps.....	4
1.4 Contributions.....	5
1.5 Outline of This Thesis.....	6
1.6 List of Publications	7
Literature Review on Fault Diagnosis Techniques for WTs	8
2.1 Traditional ML-Based Methods	8
2.1.1 Feature Extraction and Selection	8
2.1.2 Support Vector Machine.....	11
2.1.3 Decision Tree and Random Forest.....	14
2.1.4 K-Nearest Neighbor	17
2.1.5 Artificial Neural Networks.....	18
2.1.6 A Summary of Commonly Used ML Methods	21
2.2 Advanced DL-Based Methods	22
2.2.1 Convolutional Neural Network.....	22
2.2.2 Recurrent Neural Network.....	26
2.2.3 Stacked Autoencoder.....	30
2.2.4 A Summary of Commonly Used DL Methods.....	32
2.3 Generative Adversarial Network.....	33
2.3.1 Traditional Generative Adversarial Network	33
2.3.2 Auxiliary Classifier Generative Adversarial Network	35

2.3.3 Other Variants of GAN.....	37
2.3.4 A Summary of GAN.....	39
2.4 Digital Twin	39
2.5 Multi-Scale Feature Extraction	43
2.6 Summary and Link to Research Gaps	44
The Framework of a Novel Intelligent State Evaluation and Maintenance Arrangement System for WTs	46
3.1 Introduction.....	46
3.2 The Proposed Framework of the iSEMA System	48
3.3 Experiments and Results.....	55
3.3.1 Experiments and Data Description	55
3.3.2 Experiments on Offshore Data.....	57
3.3.3 Experiments on Inland Data.....	67
3.4 Discussion about the Experimental Results of the iSEMA System	71
3.5 Conclusion about the Framework of the iSEMA System	71
The Bearing Fault Diagnosis Method Based on Hybrid Framework.....	74
4.1 Introduction.....	74
4.2 The Proposed Hybrid Model.....	75
4.3 Experiments and Results.....	80
4.3.1 Experiments on the Case Western Reserve University Data	80
4.3.2 Experiments on the Jiang Nan University Data	85
4.4 Discussion about the Experimental Results of the Hybrid Model	89
4.5 Conclusion about the Proposed Bearing Fault Diagnosis Method	89
A Novel Bearing Fault Diagnosis Method Based on Multi-Scale Auxiliary Classifier Generative Adversarial Network for Imbalanced Data	91
5.1 Introduction.....	91
5.2 The Proposed ACGAN-Based Fault Diagnosis Method.....	92
5.2.1 A Brief Introduction to the Fault Diagnosis Tasks	92
5.2.2 Proposed Multi-Scale ACGAN	92
5.3 Experiments and Results.....	98

5.3.1 Experiments on the Time-Varying Rotational Speed Bearing Data.....	98
5.3.2 Experiments on the Case Western Reserve University Bearing Data.....	109
5.4 Discussion about the Experimental Results of the Proposed Fault Diagnosis Method	119
5.5 Conclusion about the Proposed Multi-Scale ACGAN-Based Fault Diagnosis Method	120
6. Conclusions.....	121
6.1 Summary and Key Findings.....	121
6.2 Addressing the Research Gaps	122
6.3 Training Time and Computational Cost.....	122
6.4 Practical Potential of the Proposed Methods	122
6.5 Directions for Future Research	123
Reference	125

LIST OF FIGURES

Figure 1. The structure of three-blade horizontal-axis WTs. [4]	2
Figure 2. SVM for Binary Classification.....	12
Figure 3. An illustration of an optimized SVM.	12
Figure 4. An illustration of decision tree.....	15
Figure 5. An illustration of random forest.	15
Figure 6. An illustration of KNN.	17
Figure 7. Illustration of a single neural unit.....	19
Figure 8. Architecture of MLP.....	19
Figure 9. An illustration of a basic CNN model.....	23
Figure 10. The structure of RNN.....	27
Figure 11. The inner structure of the LSTM unit.	27
Figure 12. The inner structure of GRU.....	29
Figure 13. Architecture of an autoencoder.....	31
Figure 14. The structure of GAN.....	34
Figure 15. The structure of ACGAN.	36
Figure 16. An illustration of DT.....	40
Figure 17. The relationship between WT and DT, together with their respective sub-systems.....	42
Figure 18. The structure of the proposed iSEMA system.....	50
Figure 19. The unrolled structure of proposed LSTM networks.....	57
Figure 20. The RMSE of LSTM under different optimization methods.....	59
Figure 21. The RMSE of LSTM for different number of hidden units.....	60
Figure 22. The RMSE distribution of LSTM in the test data.	61
Figure 23. A comparison between the actual measured power and the predicted generated power in the offshore test dataset.	61
Figure 24. The in-time comparison chart produced by the iSEMA system of offshore WT4 under condition Case C.	67

Figure 25. A comparison between the actual measured power and the predicted generated power for the inland test dataset.	68
Figure 26. Flow chart of the proposed hybrid model	76
Figure 27. The structure of 1DCNN.	77
Figure 28. Distribution of raw data.	78
Figure 29. Extracted features distribution after Conv1.	78
Figure 30. Extracted features distribution after Conv2.	79
Figure 31. Extracted features distribution after Conv3.	79
Figure 32. Extracted features distribution after the fully connected layer.	80
Figure 33. Presentation of raw data in the CWRU dataset.	81
Figure 34. Flow chart of the traditional ML model.	82
Figure 35. Presentation of raw data in the JNU dataset.	86
Figure 36. The redesigned structure of the generator.	93
Figure 37. The redesigned structure of the discriminator.	94
Figure 38. The flowchart of proposed fault diagnosis method.	97
Figure 39. Presentation of raw data in TVRS dataset.	99
Figure 40. The training process of GAN on TVRS bearing data.	101
Figure 41. The training process of ACGAN on TVRS bearing data.	103
Figure 42. The training process of Multi-Scale ACGAN on TVRS bearing data.	105
Figure 43. The fault diagnosis accuracy on TVRS bearing data with different methods for the four fault durations: 1 s, 2 s, 3 s and 4 s.	107
Figure 44. The fault diagnosis accuracy of Multi-Scale ACGAN on TVRS bearing data with different activation functions for the four fault durations: 1 s, 2 s, 3 s and 4 s.	108
Figure 45. The training process of GAN on CWRU bearing data.	111
Figure 46. The training process of ACGAN on CWRU bearing data.	113
Figure 47. The training process of Multi-Scale ACGAN on CWRU bearing data.	115
Figure 48. The fault diagnosis accuracy on CWRU bearing data with different methods for the four fault durations: 1 s, 2 s, 3 s and 4 s.	117
Figure 49. The fault diagnosis accuracy of Multi-Scale ACGAN on CWRU bearing data with different activation functions for the four fault durations: 1 s, 2 s, 3 s and 4 s.	118

LIST OF TABLES

Table 1. A summary of different types of features.....	10
Table 2. A summary of commonly used ML methods.	21
Table 3. A summary of commonly used DL methods.	32
Table 4. A summary of GAN and the commonly used variants.	39
Table 5. The algorithm of diagnosis system.	54
Table 6. The changes of Dmax with different RoAP2MP and PSDP being fixed at 50% for the offshore WT3 training data (unit: day).	62
Table 7. The change of Dmax with different PSDP and RoAP2MP being fixed at 91% for the offshore WT3 training data (unit: day).	64
Table 8. Days required to determine sub-healthy state of WT3 for different RoAP2MP for the offshore validation dataset (Unit: Day).....	64
Table 9. Days required to distinguish sub-healthy state of WT3 for different RoAP2MP under condition Case C in the offshore test dataset.....	65
Table 10. Days required to distinguish sub-healthy state of WT4 for different RoAP2MP under condition Case C.	66
Table 11. The change of Dmax with different PSDP and RoAP2MP being fixed at 91% in the inland WT1 training data.	68
Table 12. Days required to determine sub-healthy state of WT1 for different RoAP2MP in the inland validation dataset (Unit: Day).....	69
Table 13. Days required to distinguish sub-healthy state of WT1 for different RoAP2MP under the third condition case on the test dataset.....	70
Table 14. Days required to distinguish sub-healthy state for different RoAP2MP on the whole data of WT2 under the third condition case.....	70
Table 15. Parameters of 1DCNN.....	77
Table 16. Data description of the CWRU bearing data.....	81
Table 17. Explained percentage of 8 components on CWRU bearing data.....	83
Table 18. Experimental results A on CWRU bearing data.....	83

Table 19. Experimental results B on CWRU bearing data.	84
Table 20. Data description of the JNU bearing data.....	87
Table 21. Explained percentage of 8 components on JNU bearing data.....	87
Table 22. Experimental results A on the JNU bearing data.	88
Table 23. Experimental results B on the JNU bearing data.	88
Table 24. Details of generator setting.	96
Table 25. Details of discriminator setting.....	96
Table 26. Details of TVRS bearing data.....	99
Table 27. The details of three activation functions.....	108
Table 28. Details of CWRU bearing data.....	110

Abbreviations

iSEMA	Intelligent State Evaluation and Maintenance Arrangement
DL	Deep Learning
ML	Machine Learning
GAN	Generative Adversarial Network
ACGAN	Auxiliary Classifier Generative Adversarial Network
WT	Wind Turbine
DT	Digital Twin
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
SRU	Simple Recurrent Unit
RF	Random Forest
SVM	Support Vector Machine
PSO	Particle Swarm Optimization
SFLA	Shuffled Frog Leaping Algorithm
SSA	Sparrow Search Algorithm
KNN	K-Nearest Neighbor
FC	Fully Connected
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
BPNN	Backpropagation Neural Network

RBFNN	Radial Basis Function Neural Network
AE	Autoencoder
SAE	Stacked Autoencoder
SDAE	Stacked Denoising Autoencoder
ReLU	Rectified Linear Unit
RMS	Root Mean Square
RMSE	Root Mean Square Error
AR	Autoregressive
IF	Impulse Factor
MF	Margin Factor
DCGAN	Deep Convolutional Generative Adversarial Network
CycleGAN	Cycle-Consistent Generative Adversarial Network
WGAN	Wasserstein Generative Adversarial Network
REF	Roller Element Fault
OF	Out-race Fault
IF	Inner-race Fault
CWRU	Case Western Reserve University
JNU	Jiang Nan University
TVRS	Time-Varying Rotational Speed
WPD	Wavelet Packet Decomposition
PCA	Principal Component Analysis

Chapter 1

Introduction

1.1 Background and Motivations

In the 21st century, wind energy has quickly become one of the most promising and swiftly expanding renewable energy sources and now accounts for a quarter of all kinds of power generation [1]. With escalating global concerns regarding climate change and environmental sustainability, the need for clean and sustainable energy solutions has markedly risen. Compared to fossil fuels, the strategic importance of wind power lies in its sustainability and low environmental impact. Wind energy resources are abundant and renewable, and do not emit greenhouse gases during operation. It is a key component in achieving energy security and sustainable development goals. WTs, which transform wind's kinetic energy into electrical power, have thus become essential in advancing towards a low-carbon energy future. Advancements in WT technology have significantly increased efficiency and reduced costs, making wind power more competitive with conventional energy sources.

Therefore, many countries have incorporated wind energy into their national energy strategies, recognizing the potential of wind energy in diversifying energy supply and improving grid stability. Since the beginning of the 21st century, the total number of WT installations has been steadily increasing from 24 GW in 2001 to 1,021 GW by 2023. The installation of onshore WTs has experienced substantial growth, increasing from 24 GW in 2001 to 946 GW by 2023, capturing a predominant share of the market. At the same time, the installation of offshore WTs, starting from 0 GW in 2001, reached 75 GW by 2023 [2]. While the installation of offshore WTs has been gradually increasing, it remains comparatively smaller than that of onshore installation. According to [2], the growth of offshore installation accelerated notably after 2016, particularly evident from 2021 to 2023, when its expansion became more pronounced.

Modern WTs are typically three-blade horizontal-axis WTs, comprising several key components,

including the hub, blades, nacelle, gearbox, generator, tower, foundation, and control systems [3]. The structure of three-blade horizontal-axis WTs is shown in Figure 1 [4]. These components work in concert to capture wind energy and convert it into electricity. The blades capture the wind's kinetic energy and transfer it to the gearbox and generator located within the nacelle, thereby generating electricity. The produced electricity is then transmitted through the tower to the power grid. It is worth noting that, there is another popular type of WTs called direct drive WTs which converts rotor rotation to electrical power directly without the use of a gearbox [5].

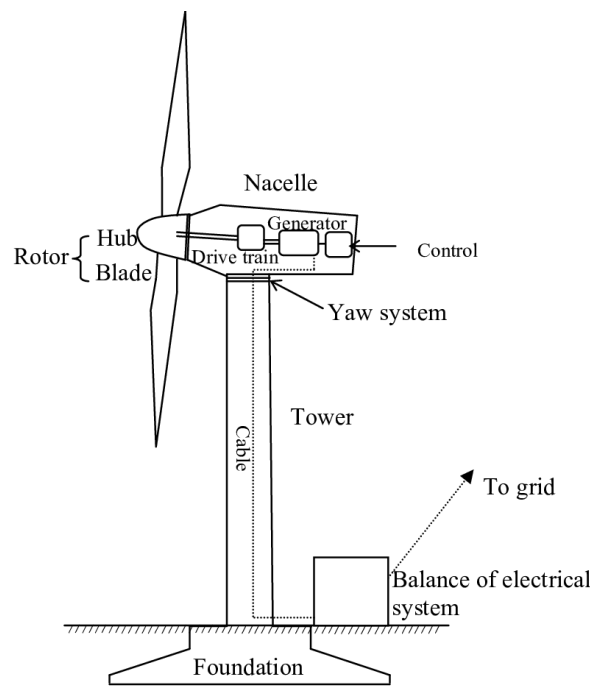


Figure 1. The structure of three-blade horizontal-axis WTs. [4]

In the past decade, there has been great research interest in the field of WT condition monitoring and fault diagnosis. Effective condition monitoring systems or methods are crucial for the reliable and efficient operation of WTs, as they help detect and diagnose potential faults early, thereby reducing downtime and maintenance costs. Vibration analysis, oil particle analysis, and acoustic emission monitoring techniques are commonly used for condition monitoring [6]. For part of the old WTs, they are about to reach the end of life; for the new WTs, their rated power and physical size are getting larger and larger [2]. Therefore, whether for old or new WTs, it is increasingly critical to ensure the reliability and efficiency of WTs. What's more, due to the high level of noise and safety problems, WTs are typically built in remote locations away from cities, such as mountains or offshore areas. This means that, besides the problems caused by the operation, WTs may encounter

environmental challenges. According to [7], the cost of maintenance after failure is nearly three times the cost of regular maintenance. If the faults can be detected before they happen and addressed during regular maintenance, it will significantly reduce maintenance costs. Thus, it is crucial to accurately assess the condition of WTs and schedule maintenance work reasonably.

The growing reliance on wind energy as a sustainable and eco-friendly power source underscores the importance of ensuring the reliability and efficiency of WTs. Current maintenance practices often lead to significant downtime and operational inefficiencies due to delayed fault detection and the high cost of maintenance after failure. Modern WTs are increasingly complex, with larger capacities and more demanding operational conditions, exacerbating the challenges of fault detection and condition monitoring. Furthermore, their remote installation sites make maintenance even more challenging and costly.

DT technology, coupled with ML and DL methods, offers a transformative approach to these challenges by enabling real-time monitoring, predictive fault detection, optimized maintenance scheduling, and accurate fault diagnosis. Additionally, addressing the sub-healthy state—where WTs exhibit performance degradation without obvious faults—can significantly enhance operational efficiency and reduce costs. Motivated by these factors, this research aims to develop an intelligent, robust, and scalable solution for state evaluation and fault diagnosis in WTs.

1.2 Aims and Objectives

This research aims to design and develop an iSEMA system that leverages DT, ML and DL technologies to enhance the reliability, efficiency, and cost-effectiveness of WT operations. By introducing the concept of a sub-healthy state, the iSEMA system seeks to proactively identify and solve potential faults, thereby reducing downtime and improving maintenance strategies.

The objectives of this research are given as follows:

- Design a DT-based framework for real-time monitoring and predictive state evaluation of WTs.
- Incorporate long short-term memory (LSTM) to construct accurate and adaptive virtual models for WTs.
- Introduce and define the sub-healthy state of WTs, providing operators with actionable criteria for early intervention.

- Design a hybrid framework combining traditional ML with advanced DL for enhanced fault diagnosis.
- Address the imbalanced data problem in fault diagnosis using an improved Multi-Scale ACGAN.
- Validate the proposed methods through extensive experiments on real-world and simulated datasets to demonstrate their effectiveness in diverse operational scenarios.

1.3 Research Gaps

Despite significant progress in WTs condition monitoring and fault diagnosis, several key research gaps remain unaddressed:

1) **Lack of a unified and intelligent framework for real-time state evaluation and maintenance scheduling.**

Although significant progress has been made in fault detection, condition monitoring, and predictive maintenance, most studies address these areas separately. There is a lack of comprehensive frameworks that integrate real-time monitoring, sub-healthy state detection, fault diagnosis, and optimized maintenance scheduling within a unified system.

2) **Limited treatment and formalization of the sub-healthy state in WTs.**

Most existing diagnosis methods classify WT states into binary categories: “healthy” or “faulty.” In practice, however, WTs often exhibit intermediate conditions where performance deteriorates without clear faults. This “sub-healthy” state is rarely formalized or incorporated into decision making processes, leading to missed opportunities for early interventions.

3) **Heavy dependence on manual feature extraction in traditional ML-based diagnosis.**

Most ML-based fault diagnosis approaches require feature extraction and feature selection manually, which limits scalability and robustness across varying conditions.

4) **Diagnosis performance degradation under imbalanced data scenarios.**

Data-driven methods for fault diagnosis typically rely on large, labeled datasets. However, in real-world applications, labeled fault data are often scarce and highly imbalanced, particularly for rare fault types. Existing approaches frequently struggle to generate reliable models under such conditions, resulting in reduced accuracy and generalization.

These gaps highlight the limitations of current research and practice. The lack of integrated

systems means that WT health management is often fragmented, reactive, and inefficient, leading to increased downtime and maintenance costs. The inability to define sub-healthy states prevents timely intervention before full failures occur, undermining predictive maintenance efforts. Furthermore, the widespread problem of data imbalance makes existing fault diagnosis methods vulnerable to poor generalization, especially for rare but critical fault modes. These challenges call for a new generation of intelligent, data-driven systems that are both comprehensive and robust. To address these challenges, this thesis proposed a framework of the iSEMA system to better evaluate the state of WTs and arrange the maintenance work. In addition, some innovative fault diagnosis methods were designed to enhance the performance of existing methods.

1.4 Contributions

This thesis makes the following key contributions to the field of WT state evaluation and fault diagnosis:

1) Development of the iSEMA System

- Designed a novel iSEMA system based on DT and DL technologies.
- The system integrates real-time monitoring and predictive maintenance capabilities, improving WT reliability and reducing operational costs.

2) Introduction of the Sub-Healthy State Concept

- Introduced and defined the concept of the sub-healthy state for WTs, a condition where performance degradation occurs without observable faults.
- Proposed criteria for identifying this state, providing a proactive mechanism to detect and address potential faults before they escalate.

3) Application of LSTM for DT Modeling

- Utilized LSTM networks to construct the virtual model of WTs within the DT framework.
- Demonstrated that LSTM-based models effectively capture temporal dependencies and dynamic behavior for accurate state evaluation.

4) Hybrid Framework for Fault Diagnosis

- Developed a hybrid model that combines traditional ML with advanced DL.
- Automated the feature extraction process, enhancing classification accuracy and efficiency

for multi-class fault diagnosis tasks.

5) Improved Fault Diagnosis Method for Imbalanced Data

- Proposed an enhanced Multi-Scale ACGAN to address the challenge of imbalanced data problem in fault diagnosis.
- Demonstrated that the improved Multi-Scale ACGAN significantly improves the stability of the training process and fault diagnosis accuracy in scenarios with limited fault samples.

6) Comprehensive Experimental Validation

- Conducted extensive experiments on both real-world and simulated datasets, including offshore and inland WT data, as well as benchmark datasets for bearing fault diagnosis.
- Validated the proposed methods under various operational conditions, demonstrating superior performance in fault detection and diagnosis accuracy.

7) Practical Implications for Wind Farm Management

- Showcased the potential of the proposed iSEMA system to minimize downtime, optimize maintenance schedules, and boost economic returns for wind farms.
- Highlighted the role of proactive maintenance strategies in enhancing confidence in wind energy as a sustainable power source, contributing to global renewable energy goals.

1.5 Outline of This Thesis

This thesis contains six chapters which are outlined as follows:

Chapter 1 introduces the background and motivation for the research, emphasizing the growing importance of renewable energy and the challenges faced in WT condition monitoring and maintenance. The aims, objectives, and contributions of this thesis are also outlined.

Chapter 2 provides a detailed review of relevant literature, including traditional ML methods, advanced DL techniques, recent developments in GANs and DT technologies, and the advancements of multi-scale feature extraction. The chapter highlights existing gaps in the field and justifies the need for this research.

Chapter 3 presents the framework of the proposed iSEMA system. The system's design, key components, and integration of DL and DT technologies are discussed. Experimental validation using offshore and inland WT datasets is provided.

Chapter 4 describes a hybrid fault diagnosis framework combining traditional ML and advanced DL models. The framework is applied to multi-class fault classification and demonstrated to improve fault diagnosis accuracy significantly.

Chapter 5 introduces an enhanced Multi-Scale ACGAN for addressing the imbalanced data problem in fault diagnosis. Experimental results demonstrate its superior performance in various operational conditions.

Chapter 6 concludes the thesis by summarizing the key findings and discussing potential future research directions.

1.6 List of Publications

- Wang, Guoliang and Hua-Liang Wei. "A Hybrid 1DCNN-SVM Model for Bearing Fault Recognition and Classification," 2022 27th International Conference on Automation and Computing (ICAC), Bristol, United Kingdom, 2022, pp. 1-6, doi: 10.1109/ICAC55051.2022.9911115.
- Wang, Guoliang, Hua-Liang Wei, and Zhao-Hua Liu. "An Intelligent State Evaluation and Maintenance Arrangement System for Wind Turbines Based on Digital Twin," Vol. 1, no. 4, Academia Engineering, 2024. <https://doi.org/10.20935/AcadEng7391>.
- Wang, Guoliang, Hua-Liang Wei, and Zhao-Hua Liu. "A Novel Bearing Fault Diagnosis Method Based on Multi-Scale Auxiliary Classifier Generative Adversarial Networks for Imbalanced Data." (Submitted)

Chapter 2

Literature Review on Fault Diagnosis Techniques for WTs

WTs operate in complex environments and are prone to various mechanical and electrical faults, which may lead to costly downtime if not detected early. Efficient fault diagnosis is therefore essential for improving operational reliability and reducing maintenance costs. This chapter reviewed the key techniques for fault diagnosis of WTs, with a focus on ML, DL, DT, and multi-scale feature extraction methods. It highlights how these techniques evolve and how they contribute to the development of the proposed iSEMA system and other innovative fault diagnosis approaches presented in this thesis.

2.1 Traditional ML-Based Methods

ML-based fault diagnosis tries to describe the relationship between monitoring data and the healthy state of the system of interest using the extracted features from the data. ML-based methods can release the manual labor to some degree and increase the diagnosis accuracy. The procedure of ML-based methods mainly consists of two parts: feature extraction and selection, and recognition and classification.

2.1.1 Feature Extraction and Selection

Feature extraction is a significant step for most fault detection and diagnosis tasks, where useful features are selected or extracted from the raw data to describe the right state of the system or process of interest (e.g., a chemical control process or an industrial machines). The feature extraction process is usually realized by converting the raw data from a high dimensional space to a low dimensional space, where the key characteristics of the data are maintained. In the meantime, such a process can also reduce the effect of noise and correlation in the raw data; models established based on the new data in the lower dimensional space will usually have better performance in terms of either accuracy

or explainability [8]. Features that are commonly used in fault detection and diagnosis can be divided into four categories represented in different format and defined in different domains, namely, time-domain features, frequency-domain features, time-frequency domain features, and other features [9].

a) Time-domain features

Mean, root mean square (RMS), variance, standard deviation, skewness, kurtosis, etc. are commonly used statistical features in the time-domain. They are easy to compute and can intuitively reflect the characteristics of raw data. Other time-domain features, such as the coefficients of autoregressive (AR) models of dynamics processes [10], impulse factor (IF), and margin factor (MF) are also used for fault diagnosis.

b) Frequency-domain features

Frequency-domain features are important and commonly used to characterize the system state. The most convenient and commonly used method for obtaining frequency-domain features is the fast Fourier transform through which one can obtain the mean frequency, median frequency, center frequency, root variance frequency, and other features. In addition, spectral entropy, spectral Shannon entropy, and spectral kurtosis are also applied as the frequency-domain features for fault diagnosis. The frequency-domain features contain some characteristics that cannot be easily found in the time-domain.

c) Time-frequency domain features

Many signals of interest are time-varying or non-stationary, for which time-frequency features would be more effective for fault detection and diagnosis tasks. A variety of signal processing methods are available for generating time-frequency domain features [11], such as short-time Fourier transform, wavelet transform, and Wigner-Ville distribution. Note that most of these methods transform a one-dimensional signal to two-dimensional space determined by two coordinates: time and frequency.

d) Other features

In addition to the features mentioned above, many other types of features are also considered in the field of fault detection and diagnosis. It is known that phase-space similarity (or dissimilarity) measurements can be used to characterize signals in fractal or fractional dimensional space. Several measurement methods have been proposed in this field, for example, the fractal dimension [12] which can describe the characteristic of complexity; correlation dimension [13] which can describe the

characteristic of self-similarity; approximate entropy [14] which can be used for describing the characteristic of regularity.

Other methods, e.g., singular value decomposition [15], piecewise aggregate approximation [16] and Kolmogorov-Smirnov test [17], have also been used to produce time-domain features for fault detection.

Feature extraction is a significant step in the process of ML-based fault diagnosis; it usually highly relies on expert knowledge. A brief summary of the features mentioned above is presented in Table 1.

Table 1. A summary of different types of features.

Types of features	Advantages	Disadvantages
Time-domain features	Easy to compute; intuitive and suitable for stationary signals	Poor performance on non-stationary or noisy signals
Frequency-domain features	Reveal hidden frequency patterns; useful for identifying periodic faults	Poor performance on non-stationary signals
Time-frequency domain features	Handle non-stationary signals; capture time-localized features	High computational cost; require careful parameter tuning

After the feature extraction step, the resulting features may still contain irrelevant or redundant information which may negatively affect the performance of the classifier. To mitigate or overcome this issue, a feature selection procedure can be performed to improve the performance and generalization ability [18]. There are mainly three types of feature selection methods, namely, filters, wrapper, and embedded methods [19, 20]. The last step of ML-based fault diagnosis methods is to design a suitable classifier to describe the relationship between the monitoring data and the health state of the system of interest (e.g. an industrial machine or equipment) based on the extracted and/or selected features. There are many ML models that can be used to realize the classification task, for example, support vector machine (SVM), K-Nearest Neighbors (KNN), decision tree, artificial neural network (ANN), and Bayes theorem [21].

2.1.2 Support Vector Machine

SVM is a powerful supervised learning method for classification and has been widely used for fault diagnosis [22] and condition monitoring [8]. SVM was initially designed to solve binary classification problems as shown in Figure 2. It can be observed that a set of boundaries which are usually called hyperplanes can separate these data into two classes. From an optimization perspective, the goal is to find the optimal hyperplane that can make the margin as large as possible as shown in Figure 3. Let the hyperplane be defined as $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$, the data points in the input space be denoted by $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the i -th input feature vector, l is the number of data points, and the labels be $y_i \in \{-1, 1\}$, then the solution to this optimization problem can be formulated as:

$$\begin{aligned} & \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \\ & s. t. y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, 2, \dots, l \end{aligned} \quad (1)$$

By applying the method of Lagrange multipliers to (1), and letting the Lagrange multipliers be denoted by α_i , then the solution for this optimization problem can be rearranged as follows:

$$\begin{aligned} & \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j), \\ & s. t. \sum_{i=1}^l y_i \alpha_i = 0, \alpha \geq 0 \quad \forall i = 1, 2, \dots, l \end{aligned} \quad (2)$$

where the indices i and j run from 1 to l and term $(\mathbf{x}_i \cdot \mathbf{x}_j)$ denotes the inner product of feature vectors. The parameter $\mathbf{w}$ can be represented as (3):

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \quad (3)$$

Thus, the optimal hyperplane is:

$$f(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}) + b \quad (4)$$

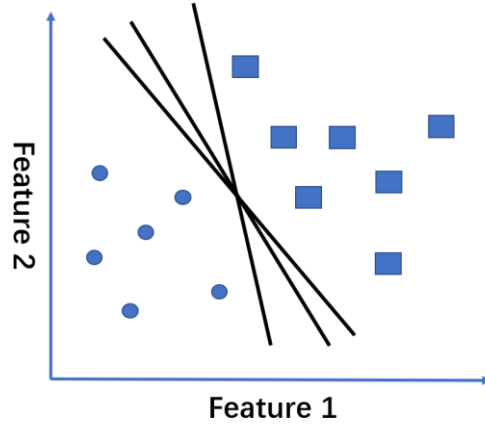


Figure 2. SVM for Binary Classification.

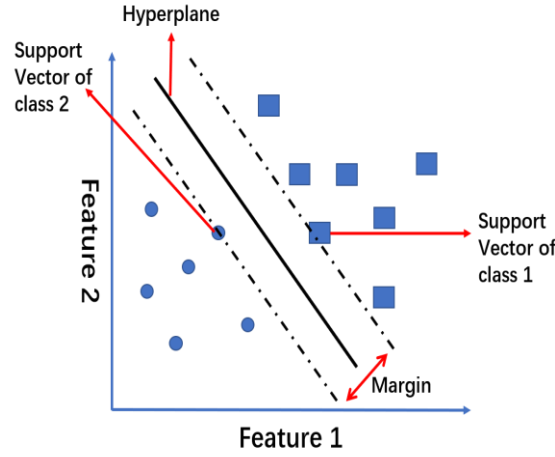


Figure 3. An illustration of an optimized SVM.

In the above SVM model, data points in the input space are usually mapped into a high-dimensional feature space by a mapping function ϕ . If the data points in the input space are linearly distributed, the solution of this optimization problem is quite similar to a binary classification problem. For nonlinearly distributed data samples, a nonlinear kernel function $K(x_i, y_i)$ can be used to solve the classification problem. The kernel function can be described as:

$$K(x_i, y_i) = \phi(x_i) \cdot \phi(y_i) \quad (5)$$

In practice, the distribution of data samples does not always exactly follow a specified law (e.g., linear relationship); for such a case, SVM may generate wrong classification results, because there are some points that are misjudged. To overcome such an issue, a relaxation variable $\xi_i (\xi_i > 0)$ and the punishment factor C will be introduced to formula (1), such that the optimization problem becomes:

$$\begin{aligned} \min \phi(\mathbf{w}) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) &\geq 1 - \xi_i, \quad \xi_i \geq 0, \forall i = 1, 2, \dots, l \end{aligned} \quad (6)$$

Similar to (2), the optimization problem can be converted into a quadratic form:

$$\begin{aligned} \min L(\alpha) &= \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^l \alpha_i \\ \text{s.t. } \sum_{i=1}^l y_i \alpha_i &= 0, \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (7)$$

Note that the decision function of this optimization problem is given by:

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i \cdot \mathbf{x}_j) + b \right] \quad (8)$$

In real applications, the selection of kernel function can significantly impact the performance of SVM. There are some commonly used kernel functions, such as linear kernel functions, polynomial kernel functions, radial basis kernel functions (Gaussian kernel), and sigmoid kernel functions. Among which, the radial basis kernel shows general good properties, e.g., powerful nonlinear mapping ability, fast convergence speed, and the involvement of fewer parameters [23]. The expression of the radial basis kernel function is:

$$K(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2g^2} \right) \quad (9)$$

where g is the kernel function parameter.

The selection of the punishment factor C and kernel function parameter g can significantly influence the performance and generalization ability of the SVM model. To improve SVM model performance, many parameter optimization algorithms have been proposed. In [24], a particle swarm optimization (PSO) algorithm was used to optimize the SVM model. What's more, multi-domain sensitive features were extracted as the input data. The experiment showed that the proposed model can achieve a better performance than models with single domain features. In [25], principal component analysis was used to extract features, then a grid search (GS) algorithm was chosen to optimize the model. The result showed that this GS-SVM model can achieve very high precision when solving fault diagnosis problems. In [26], a hybrid SVM model was designed by combining a 1DCNN with an improved novel particle swarm optimization algorithm. In [27], a Shuffled Frog

Leaping Algorithm (SFLA) was used for the parameter tuning, then the SFLA-SVM was compared with the PSO-SVM. The results showed that the SFLA-SVM model had a better performance than the traditional PSO-SVM. In [23], a novel optimization algorithm called sparrow search algorithm (SSA) was used for parameter tuning and the SSA-SVM was compared with many other SVM models. The experimental results showed that the novel SSA-SVM model can significantly increase the accuracy of fault diagnosis. In addition to various optimization algorithms, there are also a number of variants of SVM, more details can be found in [28].

SVM provides a powerful tool for solving classification problems, especially binary classification problems with small clusters. SVM has some advantages compared with other ML models, for example, it does not need a large number of sample data to train the model, the training time can be very short, and it usually has good generalization ability. However, when SVM deals with large data, a few of its obvious shortcomings must be borne in mind. First, large data may cause a huge computation cost which will slow down the training process. Second, the selection of the punishment factor and kernel function will greatly influence the performance of SVM. Third, the basic model of SVM was designed for binary classification problem but in practice, the sensitive features may have more than two classes. Thus, a complicated structure needs to be considered to solve the multi-classification problems, such as one against all, one against one, directed acyclic graph and hierarchical SVM [29].

2.1.3 Decision Tree and Random Forest

As shown in Figure 4, the top node represents the root of an inverse tree, where data splitting begins. Nodes that do not keep splitting (without branches) are leaf nodes, denoted by rectangles; nodes that keep splitting are non-leaf nodes, denoted by ellipses (except the root node). Lines between nodes are branches, representing course of action or state of nature. The widely used splitting criteria include information gain and Gini index [30].

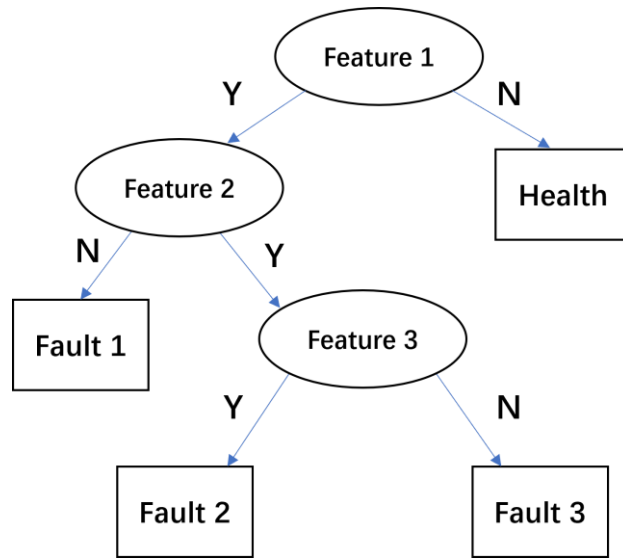


Figure 4. An illustration of decision tree.

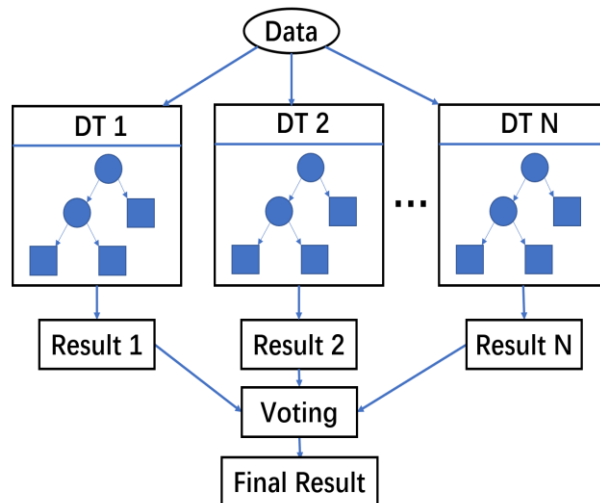


Figure 5. An illustration of random forest.

Decision tree has the following attractive properties. First, the structure of decision tree is quite intuitive and graphical which makes it easy to understand and explain. Because of this, decision tree has some characteristics of “white box” model [31, 32]. Unlike deep neural networks, decision trees generate explicit if–then rules that can be directly traced from inputs to outputs. Each node represents a decision based on a specific feature threshold, which makes the entire model structure transparent and easy to understand. However, it should be noted that such interpretability does not imply that decision trees incorporate any physical principles or domain knowledge. In this context, the term “white-box” refers purely to algorithmic transparency rather than physically grounded modeling. Second, decision tree can solve diagnosis problems even if some input data are lost. Third, decision

tree can deal with many types of data, like numerical data, and it can reach high accuracy with a faster training speed compared with many other algorithms. However, decision tree has some drawbacks. First, decision tree is easy to get overfitting. Second, if the sensitive features are highly correlated, the accuracy of decision tree can be significantly damaged [30].

As a popular supervised learning method, decision tree has been applied to address fault classification problems [1]. For example, in [33], one of the decision tree methods called classification and regression tree is combined with the adaptive neuro-fuzz inference system to fault diagnosis of the induction motors.

RF, consisting of multiple, stand-alone decision trees, is an ensemble learning algorithm, whose structure is shown in Figure 5. Data are initially divided into several subsets, each of which is then used to train an individual decision tree model. The final result will be determined by voting for all the individual decision tree models. Each decision tree model's accuracy will influence the RF model's accuracy. Generally, RF will have a better performance than a single decision tree and the problem of overfitting will also be solved, but RF usually has to train many individual decision trees and use the results generated by all the individual models to form the final result. When dealing with massive data, RF may suffer from the issue of low training speed and repeated voting. To solve this problem, in [34], a novel algorithm called Spark and improved RF was proposed. By constructing a decision tree similarity matrix, Spark and improved RF can not only improve the performance of the model but also increase the model training speed even when dealing with massive data.

In practice, it is significant to know the internal structure of a model and explain the classification results. In doing so, a good way is to take advantage of the good interpretability of decision trees and combine with other black box models to construct powerful hybrid models in recent years. For example, a deep convolutional neural network (CNN) was combined with the decision tree model in [35]. This hybrid model can not only achieve high accuracy of classification but also have good interpretability for the result. Another interesting model is called the multivariate decision trees [31]. The main difference between multivariate decision trees from the traditional decision trees is that it allows the combinations of different features during the data splitting process. More details about multivariate decision trees can be found in an insightful review [31].

2.1.4 K-Nearest Neighbor

KNN is one of the simplest classification algorithms with good efficiency for many practical applications [36]. The basic process of KNN is shown in Figure 6. The point denoted by blue circle dot at the center represents the test sample, the points denoted by squares and triangles are the training data. There are two feature classes in this example. The core idea of KNN is to search for K nearest data points around the test sample using a distance function. For example, in Figure 6, if $K = 3$, there are 2 square data points and 1 triangle point within the black circle, so the test sample is treated to be in the square class. If $K = 7$, then the test sample will be treated as the triangle class because there are 4 triangle points and only 3 square points within the red circle. From this example, it can be seen that the selection of K will directly influence the performance of the KNN model.

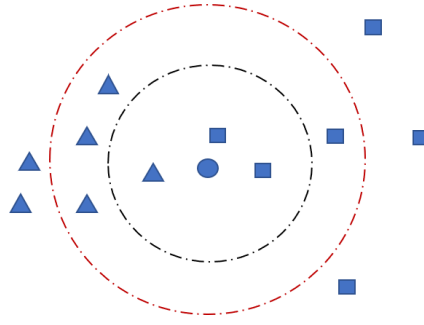


Figure 6. An illustration of KNN.

KNN is simple and powerful for solving pattern recognition problem [37] but it has some shortcomings. First, the selection of K is crucial; an unsuitable selection of K can highly influence the performance of KNN. Second, for a large dataset, KNN suffers from high computational cost for calculating the distance between the test and training samples; this also results in the requirement for high memory. Third, there are many choices of distance functions, the best choice for a specific problem may not be straightforward. As for the last drawback, a series of comparative experiments among a large number of distance functions were carried out in [36] and it concluded that there are no distance functions that can be suitable for all types of datasets. More details about the effects of distance functions on the performance of KNN can be found in [36].

Many applications of KNN in fault diagnosis have been reported in the literature. A standardized KNN was proposed for multimode fault detection and diagnosis in [38] which showed a better

performance compared with the traditional KNN method. In [39], a new KNN based fault detection method was developed using a weighted distance scheme. This new method mitigates the effect of multimode processes on KNN models and improves the performance for fault diagnosis. In [40], an enhanced KNN model was presented for the fault diagnosis of rotating machinery. This enhanced KNN model can realize the feature extraction and classification process automatically to some degree which is more intelligent compared with some traditional ML-based methods. Comparative experiments were conducted between KNN and the modified KNN in [41]; it showed that by introducing the validity and weighted vote computation procedure, the modified KNN showed a better performance of classification than the traditional KNN.

2.1.5 Artificial Neural Networks

The architecture of a typical ANN is like a human neuron information transmission model. Due to the powerful self-learning ability, an ANN has been widely used for intelligent fault diagnosis and many different types of ANNs have been developed. Multilayer perceptron (MLP) is the basic model of ANNs and is also the most widely used structure among all the ANNs [42], the illustration of a single neural unit and architecture of MLP is shown in Figure 7 and Figure 8, separately, where $x_1, x_2, \dots, x_l$ represent the inputs (features or variables), $w_1, w_2, \dots, w_l$ are the corresponding weights, b is the bias added to the neuron. After the computation, the result z is sent to an activation function $f(\bullet)$ to generate the final output. For an MLP, generally, there are three layers, the input layer, which accepts the input feature information, the hidden layer, which is responsible for the linear and non-linear calculation, and the output layer, which will send the output. MLP is a typical feedforward neural network, which is simple and flexible and has been widely used for fault diagnosis [43].

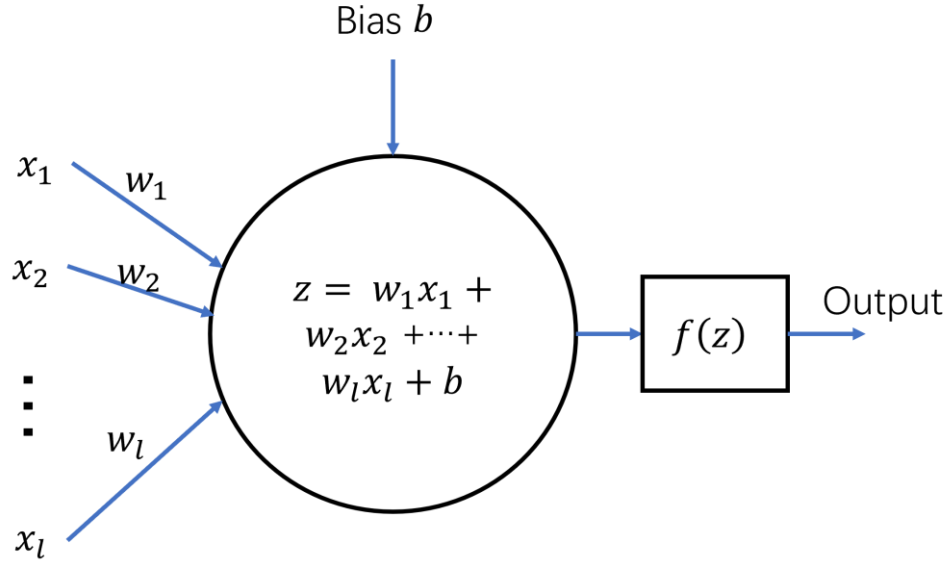


Figure 7. Illustration of a single neural unit.

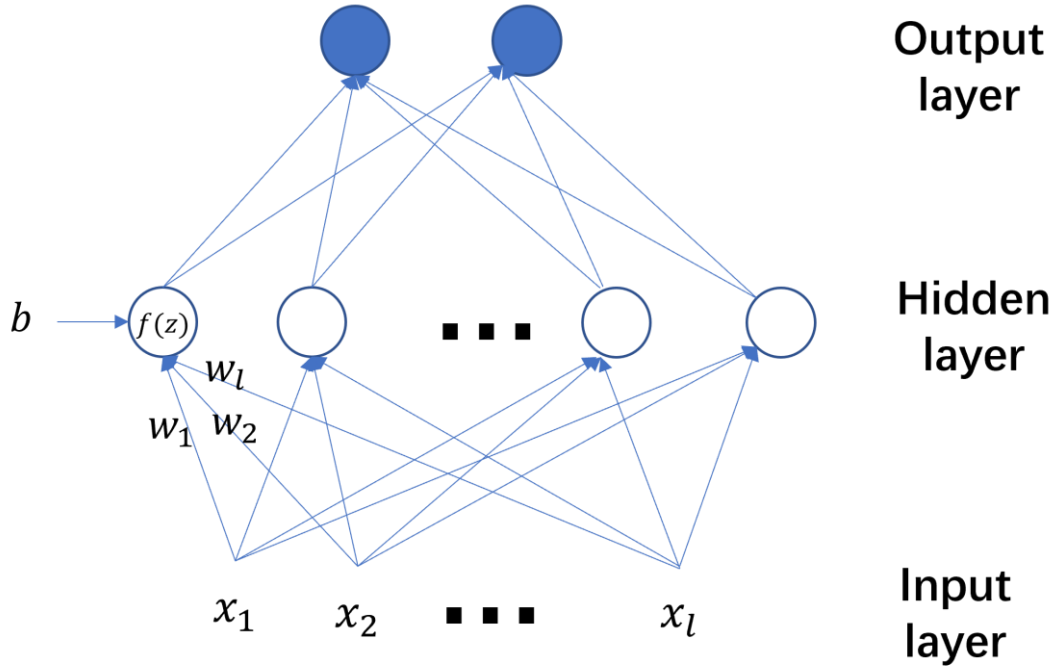


Figure 8. Architecture of MLP.

Another typical feedforward neural network is backpropagation neural network (BPNN), which can be seen as a variant of MLP based on the backpropagation algorithm. The main difference of BPNN is that it uses the gradient descent method to tune the parameters of the neural network during the backpropagation. The feedforward process produces a predicted output, and the optimization procedure is to minimize the error between the predicted output and the real output. More details on BPNN can be found in [42]. The structure of BPNN is simple: it only has three layers but in theory can well approximate any continuous functions if the hidden layer contains a sufficiently large

number of neurons. In addition, it is easy to train a BPNN and the neural network structure makes it have self-adjustment ability to some degree, so that even an error occurs in a neuron unit, the network can still generate good final output [29]. However, BPNN has its own limitations, for example, although it is easy to train, the BP algorithm itself converges slowly, and it may suffer from overfitting problem.

Radial basis function neural network (RBFNN) is another type of ANN, where the activation function is chosen to be a radial basis function, which can be described by the following formula:

$$\phi(\mathbf{x}) = \phi(\|\mathbf{x}\|) \text{ or } \phi(\mathbf{x}, \mathbf{c}) = \phi(\|\mathbf{x} - \mathbf{c}\|) \quad (10)$$

where $\mathbf{x} \in \mathbb{R}^d$ is the input feature vector, and $\mathbf{c} \in \mathbb{R}^d$ is the center vector of the radial basis function. Both are vectors in the same input space. As shown in (10), radial basis function is a kind of function that focuses on the distance from any point in the space to the center point $\mathbf{c}$.

In the application of RBFNN, it is important to confirm the number and location of the neurons in the hidden layer according to center $\mathbf{c}$. What's more, RBFNN needs to calculate the distance between the input points and hidden layer neurons, and this makes the calculation of RBFNN more complex than that of BPNN. However, the structure of RBFNN is simple and it has a powerful nonlinear fitting capability and generalization capability. In addition, it has a faster convergence speed, so the training process is highly improved, thus, RBFNN has been widely used for fault diagnosis [29].

In comparison with many other traditional ML methods, ANN has better self-learning capability and better performance for fault diagnosis and therefore has been extensively developed used for fault diagnosis. Besides these three typical neural networks, MLP, BPNN and RBFNN, there are also many other types of neural networks, for example, extreme learning machine [44], self-organizing map [45], and adaptive resonance theory [46]. In addition, compared with the traditional ML methods, ANN has a better performance when dealing with high-dimension data. It is worth mentioning that all these ANNs need human interference in the process of feature extraction and feature selection. What's more, the number of neurons in hidden layers can highly influence the accuracy of ANN models, but there are no standard rules for choosing the best number of neurons. Fewer neurons mean poorer accuracy, more neurons can improve the performance, but more training time will be required and the model may suffer from overfitting problems.

2.1.6 A Summary of Commonly Used ML Methods

The aforementioned methods all have their own advantages and disadvantages, therefore carefully choosing a suitable one to solve a given fault diagnosis task would be very important. A brief summary of these methods is shown in Table 2.

Table 2. A summary of commonly used ML methods.

Methods	Advantages	Disadvantages
SVM	• Effective for binary classification tasks • High accuracy with small datasets [23]	• Computationally expensive on large-scale data [29] • Requires kernel function selection and parameter tuning [23] • Not naturally suitable for multi-class problems without structural extensions [29]
Decision Tree	• Clear structure to understand and explain [31] • Handles both numerical and categorical data	• Prone to overfitting due to structural instability from minor data perturbations, excessive node splitting to fit noise, and fragmented feature optimization in high-dimensional spaces [30]
KNN	• Simple to implement • Suitable for pattern recognition tasks [37]	• Sensitive to the choice of K and distance metrics [36] • High computational cost for large data [37]
ANN	• Good for modeling complex nonlinear relationships • High accuracy when tuned properly [29]	• No standard rules for hyperparameter tuning [29]

These limitations of traditional ML models motivate the hybrid structure proposed in this thesis, where DL models are utilized for automatic feature extraction, and interpretable ML models like RF are adopted for final classification, leveraging the strengths of both.

2.2 Advanced DL-Based Methods

With the development of CPU and GPU, the performance of DL, which is the branch of ML, has been highly improved and widely used in many fields, for example, object recognition, speech recognition, natural language processing, image processing, medical applications, etc. [47]. DL models usually have a much more complex structure than traditional neural network models, which means that the training process of DL models will require much more time. However, the complex structure makes DL have a more powerful self-learning capability which releases manual labor. In addition, DL models usually have better performance than traditional ML models especially when dealing with massive data. In recent years, various DL models have been proposed, for example, CNN, recurrent neural network (RNN), stacked autoencoder (SAE), and deep belief networks. Each type of neural network model has a number of variants.

2.2.1 Convolutional Neural Network

a) 2DCNN

CNN, which was initially designed to process 2D images [48], has been widely used in many areas, including image identification and classification, pattern recognition, and speech processing. It has also been used for fault diagnosis due to its powerful self-learning and feature extraction capability. The structure of a basic CNN model comprises a number of convolutional layers and pooling layers, a fully connected (FC) layer, and an output layer, as shown in Figure 9. A convolutional layer performs the convolution operations by its kernels and these kernels have two significant characteristics: sparse connection and shared weights [49]; in the same convolutional layer, there are usually multiple convolutional kernels, and these kernels only act with the local region of the previous input with the constant weights. Let $\mathbf{x} \in \mathbb{R}^d$ be a d-dimensional input data, then on the j -th feature map, the i -th convolution result can be expressed as:

$$c^i = \mathbf{x} * \mathbf{w}^j + b^j \quad (11)$$

where the symbol ‘*’ represents the convolution operation; $\mathbf{w}$ and b are the filters and basis respectively.

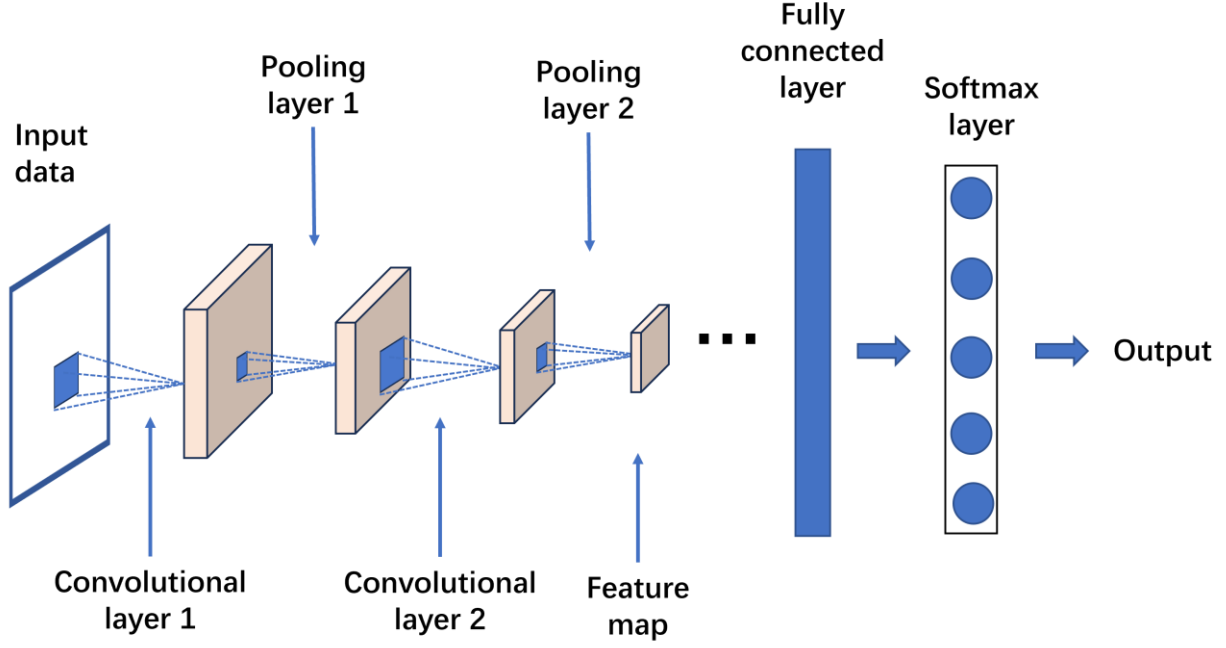


Figure 9. An illustration of a basic CNN model.

After the linear operation, the following process is the non-linear operation, in which the activation function plays a key role. There are several commonly used activation functions, such as, sigmoid function, tanh function, and rectified linear unit (ReLU) function, which are defined as follows respectively:

$$\text{sigmoid: } f(z) = \frac{1}{1 + e^{-z}} \quad (12)$$

$$\text{tanh: } f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (13)$$

$$\text{ReLU: } f(z) = \max(0, z) \quad (14)$$

Generally, the selection of activation function depends on the situation, but ReLU is widely used due to its faster computation speed and the inhibiting ability of gradient vanishing problem. However, it should be noticed that the gradient is zero when z is close to and smaller than zero. To solve this issue, an improved activation function called leaky ReLU was proposed [49].

The pooling layer is another important layer which performs the pooling operation. In the pooling operation, the dimensions of feature maps will be reduced, and the number of parameters will also become smaller. Meanwhile, during the pooling operation, the main characteristics of feature maps will be maintained. There are two types of pooling layers, namely, max pooling layers and average pooling layers, both split the feature map into several sub-maps, the difference is that a max

pooling layer chooses the maximum value of sub-maps as the output, but an average pooling layer chooses the average value of the sub-maps as the output.

After the convolutional layers and pooling layers, the extracted features go to the FC layer whose function is the same as in MLP. The extracted features will be processed by the FC layer and then sent to the output layer for classification or other tasks.

In recent years, besides the basic CNN model, many improved variants and associated approaches have been proposed. Residual network [50] was proposed to address the problems of gradient vanishing or gradient exploding by adding the extra residual blocks. A densely connected convolutional network was introduced in [51], where the dense connection blocks were added to the CNN model. More specially, any layer will receive information not only from the previous layer but also from all other previous layers. GAN [52], which consists of a generative model and a discriminative model, is getting more and more attention in the field of DL due to its excellent data generation capability. In addition, there are also some useful techniques that can be used in the process of model training. Batch normalization [53] is a common technique and has been widely used in the training of CNN models due to its ability to reduce data covariance shift, and it can further promote the training process and improve the performance of the model. Dropout [54] is an important approach for improving CNN model performance. By randomly dropping some neuron units in the training process, it can inhibit the overfitting problem effectively.

It is known that most existing CNNs are designed for dealing with 2D data (e.g., images). In many real applications, however, the raw data are recorded and represented as a set of time series signals (such as vibrations signals), which cannot be directly processed or modelled with existing CNNs. Therefore, in recent years extensive effort has been made to develop methods that can handle 1D signals. One approach is to convert the 1D raw data into a 2D format with the main characteristics being maintained. The resulting 2D data can then be stacked together into a 2D matrix. In [55] and [56], the raw data was transformed into a 2D matrix, then the matrix was sent to 2DCNN models to detect the faults of rolling bearings. Another transformation method is to convert raw data into pixel images which are then processed with a CNN model. For example, a novel model called CarNet was proposed in [57], where the 1D time sequence was first converted to 2D images which were used to train a CarNet model. In [58], the raw vibration data were transformed into color image data using an image conversion approach to train a CNN model. In addition to the aforementioned two

transformation methods, other methods developed based on signal processing approaches, for example, wavelet transform and short-time Fourier transform, have also been used to convert 1D raw data to 2D representations (e.g., in time-frequency domain) which are used for CNN network model training. In [59], the raw data were used to train the CNN model separately, namely time-domain, frequency domain, and time-frequency domain. Results from existing studies (e.g., [59]) show that an CNN model trained with time-frequency domain features showed the better performance for fault diagnosis of rolling bearings than that models trained with the time-domain and frequency-domain data alone.

b) 1DCNN

The approaches discussed in the previous section can convert the raw data into a 2D format that can be processed with 2DCNN. However, these methods require manual interference in one way or another. This is not good in terms of implementation efficiency. In fact, CNN models can be more efficient if such a “1D to 2D” data transformation and processing procedure can be circumvented. To this end, 1DCNN was proposed to directly deal with 1D data.

Raw data collected from industrial machines, such as vibrations, current, acoustic emissions, and instantaneous speed, are commonly represented as time-series. Among which, vibration signals are the most commonly used information in fault diagnosis of rotating machines [60]. In [61], an adaptive overlapping CNN was proposed to process vibration signal directly. The selection of filter size and numbers for a real data set was discussed in [62] and the fault diagnosis method based on 1DCNN model was compared with other traditional ML-based methods. The experimental results showed that method based on 1DCNN model had a better performance for bearings fault diagnosis. In [63], a threshold model and a CNN model were combined for bearing fault diagnosis. By adding the threshold model, the physical meaning of the proposed method was increased. Huang et al. [64] employed 1DCNN to extract and select features automatically from vibration signals, then the features were used to train the deep decoupling CNN model for the compound fault diagnosis. In [65], a 1DCNN model was used to conduct the feature learning process from raw vibration signals directly, then the sensitive features were sent to the bidirectional long short term memory model for detecting the wear state of tools. Acoustic emission data are good at detecting an early fault of machines, especially for machines working under the condition of low speed and low frequency noise. In [66], a hybrid model consisting of a 1DCNN and a gated recurrent unit (GRU) was proposed for the

detection of the gear pitting problem. The CNN model was trained with acoustic emission data and the GRU model was trained with vibration signals. The presented hybrid model had a good performance on the diagnosis of gear pitting which can reach up to 98% accuracy. Different from vibration and acoustic emission signals, current signals are commonly used for the detection of machines driven by electricity, for example, motors. In [67], real-time motor stator current data was employed to train the 1DCNN model and another traditional classification model separately. The comparative experiments demonstrated that the 1DCNN model was better than the traditional classification model. In [68], a comparison experiment was designed to evaluate and compare the performance of 1DCNN and LSTM models. Both were trained with the motor current data under various load conditions and the results demonstrated that the CNN model was more suitable for the real-time fault diagnosis of motors.

Deep 2DCNN, due to its structure (e.g., numerous hidden layers and massive parameters), is powerful for dealing with complex projects and large amounts of data, and especially suitable for 2D signals, such as images. But if the model input is a 1D signal, e.g., a time-series, then 1DCNN will be a better choice. In addition, because of the compact structure (with 1-2 hidden layers) and fewer parameters, 1DCNN performs better than 2DCNN in training speed and is therefore more suitable for a real-time application [69].

2.2.2 Recurrent Neural Network

RNN is one important type of DL model and has been widely employed in many different fields. Compared with other DL models, the main difference is that RNN has a powerful capability of handling time series data and extracting nonlinear features [70]. Figure 10 shows the basic structure of RNN which has a special delay layer that can store previous memory [48]. Although RNN has a good performance in handling time series data, the traditional RNN suffers from gradient disappearance and gradient explosion. In addition, traditional RNNs often struggle to capture long-term dependencies due to the vanishing and exploding gradient problems [70]. To solve this problem, the gate mechanism [71] was introduced to form a special type of RNN, called LSTM, and the internal structure is shown in Figure 11.

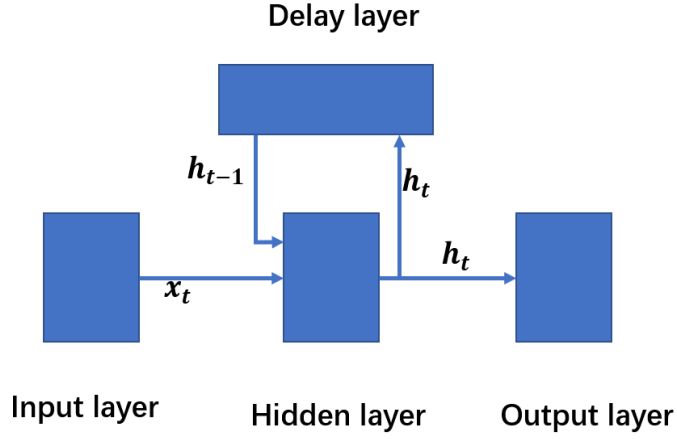


Figure 10. The structure of RNN.

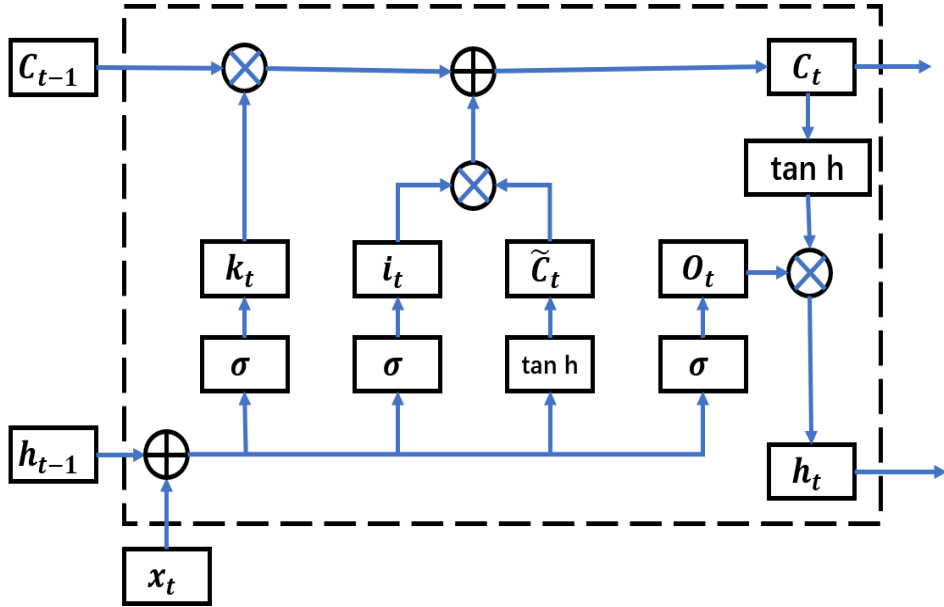


Figure 11. The inner structure of the LSTM unit.

The introduction of three gates and a memory cell state c_t into the RNN structure enables LSTM to have both long-term and short-term memory; the hidden state is known as short-term memory and the cell state is known as long-term memory [72]. The calculation of the LSTM unit can be described by the following equations:

$$\tilde{c}_t = \tanh \left(\mathbf{w}_c \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} + \mathbf{b}_c \right) \quad (15)$$

$$o_t = \sigma \left(\mathbf{w}_o \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} + \mathbf{b}_o \right) \quad (16)$$

$$i_t = \sigma \left(\mathbf{w}_i \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} + \mathbf{b}_i \right) \quad (17)$$

$$\mathbf{k}_t = \sigma \left(\mathbf{w}_k \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}_{t-1} \end{bmatrix} + \mathbf{b}_k \right) \quad (18)$$

$$\mathbf{c}_t = \mathbf{k}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (19)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (20)$$

where, $\mathbf{k}_t$, $\mathbf{i}_t$, and $\mathbf{o}_t$ represent the forget, input gate, and output gate, respectively; $\mathbf{x}_t \in \mathbb{R}^d$ is the input vector at time step t and $\mathbf{h}_{t-1} \in \mathbb{R}^h$ is the hidden state from previous time step; $\mathbf{c}_t$ and $\tilde{\mathbf{c}}_t$ are the updated cell state and the candidate cell state; $\mathbf{w}_c, \mathbf{w}_o, \mathbf{w}_i, \mathbf{w}_k \in \mathbb{R}^{h*(h+d)}$ are the weight matrices associated with cell state, output gate, input gate and forget gate, and $\mathbf{b}_c, \mathbf{b}_o, \mathbf{b}_i, \mathbf{b}_k \in \mathbb{R}^h$ are the bias vectors associated with cell state, output gate, input gate and forget gate; $\odot$ denotes the element-wise multiplication; $\sigma(\cdot)$ represents the sigmoid activation function.

Equation (15) to equation (20) mathematically describes the internal operations of the LSTM unit illustrated in Figure 11. Specifically, the figure shows the flow of input $\mathbf{x}_t$ and the hidden state of previous time step $\mathbf{h}_{t-1}$ through the network, where they are jointly transformed to compute the input gate $\mathbf{i}_t$, forget gate $\mathbf{k}_t$, output gate $\mathbf{o}_t$, and candidate cell state $\tilde{\mathbf{c}}_t$ via respective activation functions. These components are then used to update the cell state $\mathbf{c}_t$ and generate the hidden state $\mathbf{h}_t$. Thus, the figure provides a visual representation of equation (15) to equation (20) governing the LSTM dynamics.

Note that the introduction of the gate mechanism can lead to a significant increase in the training time. To solve this issue, an improved LSTM unit named GRU was proposed. The internal structure of the GRU unit is depicted in Figure 12. In the GRU unit, the forget gate and input gate are mixed as an update gate $\mathbf{z}_t$, and the cell state and hidden state are combined as a reset gate $\mathbf{r}_t$. The number of gates becomes two, which is smaller than the number of gates in LSTM, so GRU is more compact and efficient and usually takes less time to train.

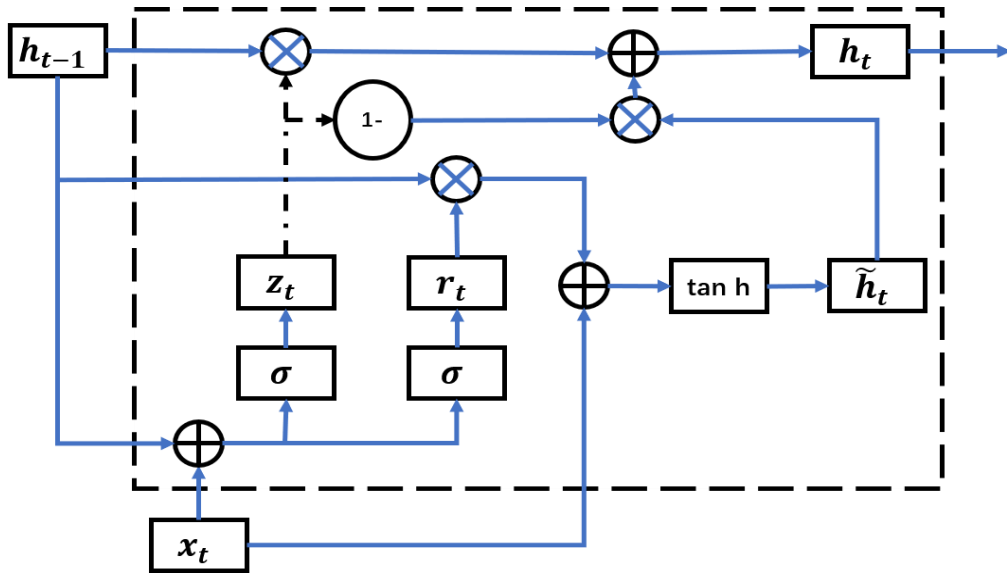


Figure 12. The inner structure of GRU.

Due to its capability of handling sequential data including time series, LSTM and GRU have been widely used in the field of fault diagnosis. In [73], the GRU model was employed to handle multi-dimensional inputs for fault diagnosis and identification of the blast furnace ironmaking process; the result demonstrated that the proposed model had a higher accuracy than other compared models. In [74], many LSTM models were trained with the vibration signals, and the model parameters were determined by using a Bayesian approach. The accuracy of the best LSTM model reached up to 93% which was superior to both the compared traditional classification methods and the compared DL-based methods. Miao et al. [75] presented a real-time fault diagnosis model based on GRU and the experimental results proved that the GRU model can realize the fault diagnosis tasks with fast speed and high accuracy. Xue et al. [76] employed LSTM to a WT system for fault diagnosis of switches; its performance was compared with that of least square SVM and conventional RNN. The experiments showed that the proposed LSTM model got the highest accuracy and best robustness performance. In [77], a CNN model was used to extract the spatial features, and a LSTM model was used to capture the sequence features. Then, these two types of features were fused for bearings fault diagnosis. The results showed that the proposed method can recognize the fault with a high accuracy even under strong noises conditions.

It is known that in traditional RNN and LSTM, the values are updated through serial computation, meaning that the update must wait for the completion of the previous computation at each step, and therefore has a great limitation in its training efficiency. To improve the training

efficiency of RNN and LSTM, Lei et al. [78] proposed an improved RNN called simple recurrent units (SRUs). The experimental results demonstrated that SRU has the same efficiency as CNN and feed-forward network in parallel computation. The internal structure of SRU can be divided into two parts, namely, light recurrence and highway network. It also has two gates, that is, forget gate and reset gate. Such a compact and light internal structure enables SRU to take less time to train than LSTM and GRU.

Besides LSTM, GRU, and SRU, there are also many other RNN based fault diagnosis methods, for example, bidirectional RNN based methods, and RNN encoder-decoder based methods. For example, in [79], a deep bi-directional LSTM model was proposed for fault detection of WT gearbox. The proposed model was compared with BPNN, SVM, CNN, and LSTM, and the results demonstrated that bi-directional LSTM showed promising performance. In [80], a fault diagnosis method based on RNN encoder-decoder was presented and compared with a CNN based method, the results showed that the proposed method outperformed the CNN based method.

2.2.3 Stacked Autoencoder

Autoencoder (AE) [81] is an unsupervised learning algorithm consisting of three layers: input layer, hidden layer, and output layer, which are shown in Figure 13. The procedure can be divided into two parts: encoder process and decoder process, where, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are the input vector, $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$ are the encoder vector in hidden layer, and $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ are the output vector. Here, $\mathbf{W}_1$ is the weight matrix to realize encoder process and $\mathbf{W}_2$ is the weight matrix to realize decoder process. The core idea of AE is to decrease the disparity between the generated output and original input.

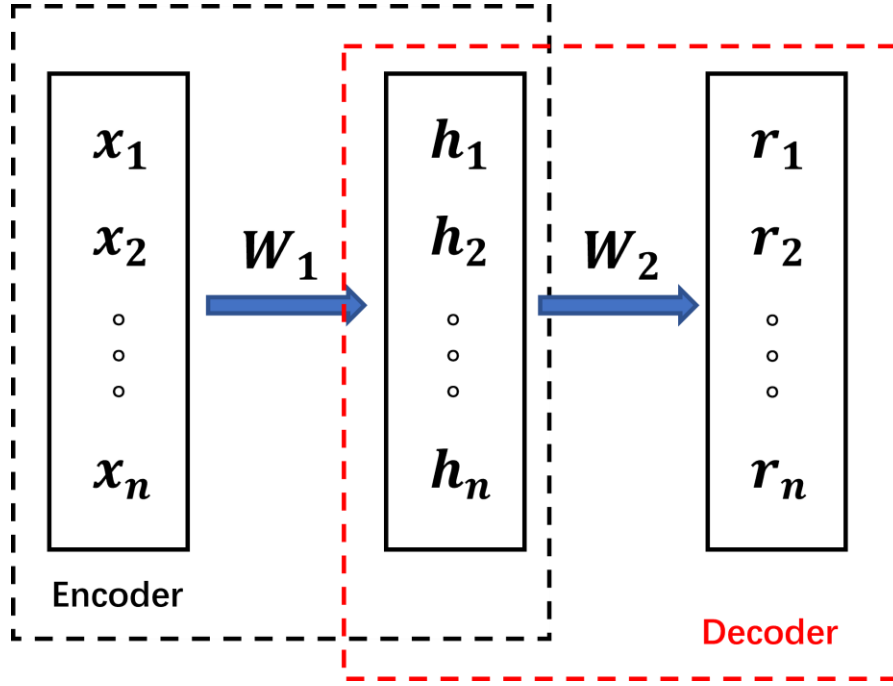


Figure 13. Architecture of an autoencoder.

To obtain more characteristics from input data, SAE is proposed [82]. After the training process for the first AE, the decoder part is removed and the encoder vector $h_1, h_2, \dots, h_n$ can be seen as the feature vector obtained from the input vector. Then, the feature vector is used as the input vector to train the second AE, and the decoder part is also removed after the training process. Therefore, repeating the process several times, the AEs are stacked together. In [83], SAE was employed to enhance the vibration signals. Then, the processed data was converted into 2D images to train a CNN model for fault diagnosis. After the enhancement process, the results showed that the training accuracy was improved to 98.89% and the test accuracy was improved to 97.25%. However, SAE may face a problem that the basic unit AE may copy the input vector without obtaining any useful characteristics of input. To address this problem, denoising AE is proposed by adding noise to AE which also increases the robustness ability [84]. Stacked denoising autoencoder (SDAE) has been applied to bearing fault diagnosis in recent years. In [85], a deep neural network structure was designed based on SDAE. The experimental results showed that the proposed method can classify the combined faults even only single fault data used as the training data. In [86], SDAE was employed for feature extraction and dimension reduction from bearings vibration signals. The experimental results indicated that methods based on SDAE performed better than other compared methods.

Besides the mentioned denoising AE, there are two variants of AE: sparse AE and contractive AE. Sparse AE is constructed by adding a sparsity penalty term to the loss function of the basic AE

which enables the sparse AE to learn more sparse and representative features of input data [87]. In [88], a fault diagnosis method based on stacked sparse AE was proposed to automatically extract sparse and representative features for rotating machinery. Contractive AE is constructed by adding a contractive penalty to loss function which makes it can learn more robust representations of input data [89]. Shen et al. [90] proposed an automatic and robust features learning methods based on the stacked contractive AE for gearbox and bearing fault diagnosis. The accuracy of the proposed method can reach up to nearly 100% which is superior to other compared methods.

2.2.4 A Summary of Commonly Used DL Methods

Different DL models have different characteristics, and a brief summary of the mentioned DL models is concluded in Table 3.

Table 3. A summary of commonly used DL methods.

Methods	Advantages	Disadvantages
CNN	- Fast training speed - Good performance	- High requirements for labeled data - May lose time dependencies of sequential data
RNN	- Good at handling sequential data - Can reflect time dependencies of input data	Slow training speed
SAE	Particularly suitable for feature extraction	AEs must be trained first before constructing the model

Compared with traditional ML-based methods, DL-based methods have some advantages. First, DL-based methods usually can extract more representative and relevant features without manual interference from massive data because of the deep structure. Secondly, DL-based methods usually have more generalization ability and accuracy than traditional ML-based methods [91]. In theory, even a very complicated problem can be addressed by well-trained DL (e.g., well trained through sufficiently large training samples) if there are enough number of layers and neurons. However, it also has some limitations. Firstly, it usually requires a large or big training dataset to achieve good

DL models. Secondly, DL models usually have more complicated structures so the training for a DL model is time consuming, with a high requirement for GPUs and CPUs. Thirdly, DL models usually have a huge number of parameters but so far there is no standard theory for the selection and tuning of the parameters. More importantly, in most cases the output of DL models cannot be interpreted and explained, and this highly influences the applications of DL-based methods in problem solving tasks where result explanation (e.g., cause-effect relationship) is highly required or desired.

So far, some of the existing ML methods for fault diagnosis have been introduced. In classical ML part, commonly used features are introduced, and they can be divided into four categories represented in different format and defined in different domains, namely, time-domain features, frequency-domain features, time-frequency domain features, and other features. Then, the commonly used ML algorithms are discussed, and the advantages and disadvantages are concluded in Table 2. Among these, decision tree (or RF) has a clear structure and can deal with multi-classification tasks, so it is a good selection for the classification tasks in this project. In DL part, the advantages, and disadvantages of CNN, RNN and AE are discussed, and the key points are concluded in Table 3.

2.3 Generative Adversarial Network

2.3.1 Traditional Generative Adversarial Network

The concept of GANs was introduced by Goodfellow et al. [92] as a novel framework for training generative models and the structure is shown in Figure 14. GANs consist of two neural networks: a generator G and a discriminator D , which are trained simultaneously through an adversarial process. Generator G aims to produce data samples that are indistinguishable from real data, while discriminator D evaluates the authenticity of the samples, distinguishing between real and generated data. The objective function of a GAN can be defined as:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_Z(z)} [\log(1 - D(G(z)))] \quad (21)$$

where x is a real data input vector, z is a random noise or latent variable vector. $G(z)$ represents the generated data. $D(x)$ is the probability that the real data input x is a real data sample as estimated by the discriminator D , $D(G(z))$ represents the probability that the generated data $G(z)$ is seen as a real data sample by the discriminator D . $E(*)$ is the expectation operator which is used to compute

the average value of a random variable under a given probability distribution. $P_{data}(x)$ is the probability distribution of the real data and $P_z(z)$ is the prior distribution of the noise z . The objective function is a min-max optimization problem involving two players: the generator G and the discriminator D . The discriminator D aims to maximize the objective function, and it tries to correctly distinguish between real data input x and generated data $G(z)$. The generator G aims to minimize the objective function, and it tries to produce fake data that are indistinguishable from real data x , fooling the discriminator. In Figure 14, the loss functions for the generator and discriminator are usually respectively defined as follows:

$$L_D = -E_{x \sim P_{data}(x)}[\log D(x)] - E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (22)$$

$$L_G = -E_{z \sim P_z(z)}[\log D(G(z))] \quad (23)$$

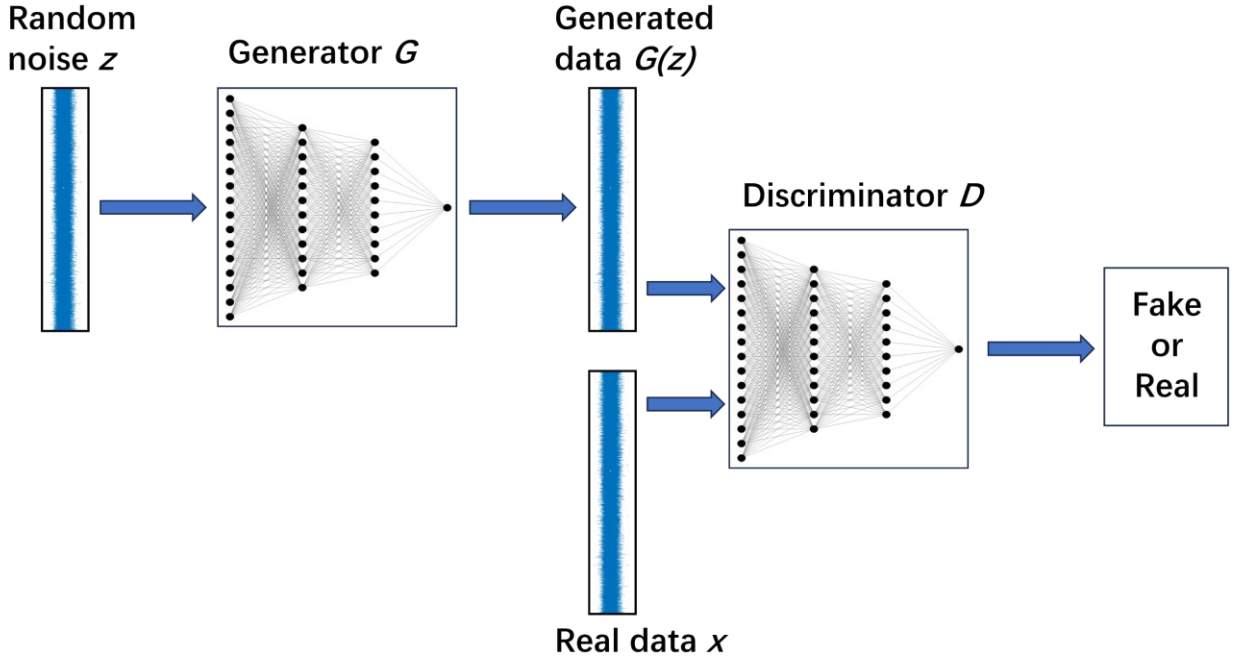


Figure 14. The structure of GAN.

Since their inception, GANs have been applied to various domains, including natural language processing [93], computer vision [94], image synthesis [95], voice synthesis [96], target recognition [97], password cracking [98], burst gravitational wave generation [99], and data-driven materials design [100]. A comprehensive review about GAN can be found in [101].

In the field of fault diagnosis, many improved methods based on GAN have been proposed. In [102], a novel gearbox fault diagnosis method is presented by combining SDAE and GAN. In [103], it was pointed out that the GAN-based fault diagnosis methods usually ignored the similarity in fault

characteristics. To address this problem, the outputs of a CNN and the envelope spectrum error of the Fault Characteristic Frequencies were used as two more correction terms to improve the performance of GAN. In [104], a fusion attention mechanism and a global feature cross module was combined into the generator and discriminator, respectively. In addition, an anti-noise module is employed to improve the anti-noise performance of GAN. These efforts have significantly improved the accuracy of bearing fault diagnosis. In [105], a bearing fault diagnosis model was designed based on GAN with the 2D spectrogram of the acoustic emission signals. The proposed method significantly improved the performance of the traditional CNN-based fault diagnosis method under low samples and data imbalance condition. In [106], an improved GAN-based fault diagnosis method was introduced to address the limitation of fault data and poor diversity of generation. The proposed method was validated on both rotor and rolling bearing data and the results demonstrated that it can significantly improve the classification accuracy. In [107], the traditional GAN structure was redesigned by incorporating the embedding network and RNN, where the loss function was defined as the associated mean squared errors for better training performance. The fault diagnosis accuracy of the proposed method was significantly improved compared to other traditional fault diagnosis methods. In [108], a new fault diagnosis method for WT was developed based on GAN and temporal CNN, where GAN was employed for feature extraction and temporal CNN was employed for classification. It was shown that the method outperformed the compared methods. In [109], a vibration data generation method was presented for rotating machinery fault diagnosis based on an improved GAN where the loss function was defined based on a multi-resolution short-time Fourier transform. In [110], a novel fault detection method designed based on GAN and DT was reported, where the GAN model was employed for virtual representation of sensors and the GAN discriminator was employed for fault detection. In [111], a novel rolling bearing fault diagnosis method based on Wasserstein GAN and SAE was proposed to address the data imbalance problem. The experimental results indicated that the proposed method was superior to the conventional GAN and the variational AE method.

2.3.2 Auxiliary Classifier Generative Adversarial Network

ACGANs were introduced as an extension of GANs that incorporates class labels into the learning process [112]. Its structure is shown in Figure 15. In the ACGAN, the input data could be a

random noise z or the label information c ; the discriminator not only distinguishes between real and generated samples but also predicts the classes of the input data. Therefore, the output of the discriminator includes two parts: 1) $P(S|X)$: the probability distribution of whether the data is real data, and 2) $P(C|X)$: the probability distribution of what the label is. This additional objective guides the generator to produce class-specific data samples.

The ACGAN objective function includes two parts: the source loss L_s and the class loss L_c . The source loss remains similar to that of traditional GANs, while the class loss ensures the correct classification of real and generated samples. The objective functions for the discriminator and generator are given by the following equations, respectively:

$$L_s = E[\log P(S = \text{real}|X_{\text{real}})] + E[\log P(S = \text{fake}|X_{\text{fake}})] \quad (24)$$

$$L_c = E[\log P(C = c|X_{\text{real}})] + E[\log P(C = c|X_{\text{fake}})] \quad (25)$$

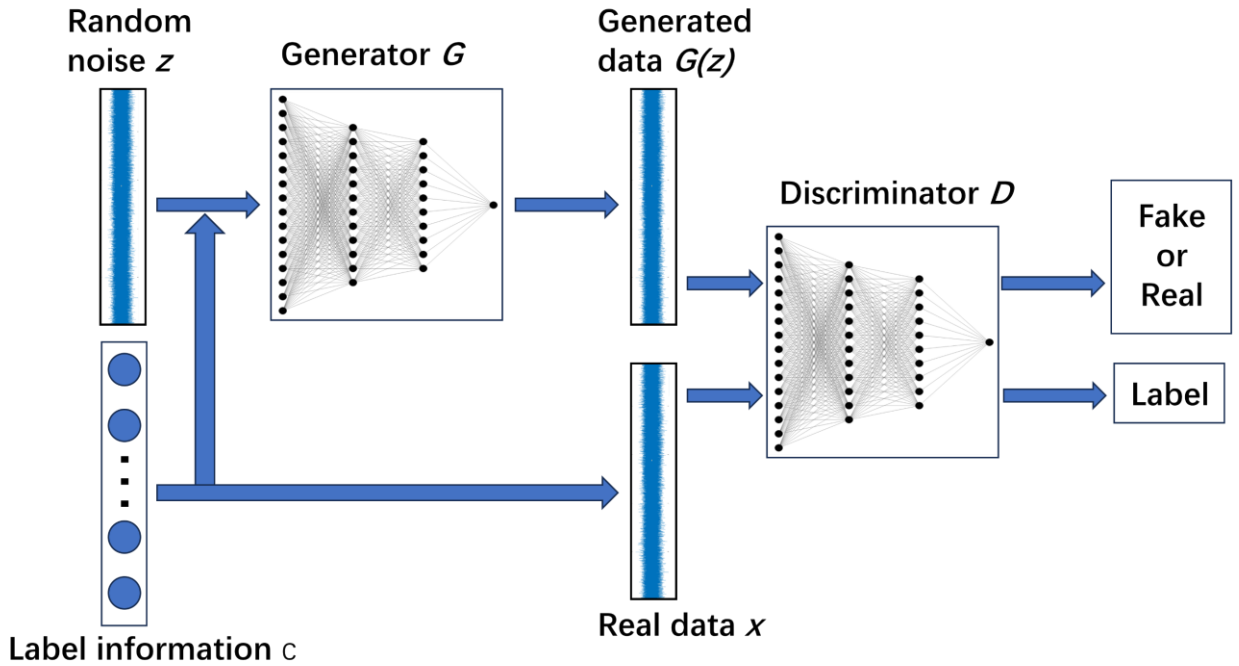


Figure 15. The structure of ACGAN.

During the training process, the discriminator is trained to maximize $L_s + L_c$, as the discriminator is trying to distinguish whether the data is real or not and make sure the correct classification. While the generator is trained to maximize $L_c - L_s$, because the generator is trying to produce fake data which can be considered as real as much as possible and can be effectively classified.

By introducing additional label information and the auxiliary classifier to the traditional GAN, ACGAN can significantly improve the quality of generated data and make the training process more stable, thereby, much research has been carried to exploit the potential of ACGAN [113]. In [114], an enhanced ACGAN was proposed to address the common problem of fault diagnosis under imbalanced data. By introducing the boundary seeking loss and sensitive learning, the proposed ACGAN performed better than other GAN-based methods. In [115], ACGAN was combined with a transformer network to carry out experiments on small motor bearing datasets. The proposed model can avoid the coupling problem which is commonly seen in traditional ACGAN and have a simpler structure. In [116], ACGAN was combined with the deep residual network to enhance the data. The experimental results demonstrated that the proposed methods can significantly improve the performance under imbalanced data. In [117] a new ACGAN framework is designed by adding an independent classifier and proposed new loss functions based on Wasserstein distance. To extract features from different time periods without using manual feature extraction, the proposed methods in [18] transformed the raw vibration data into 2D images. Experimental results demonstrated that their proposed methods could generate data with higher qualities and improve the performance of fault diagnosis. In [118], a bearing fault diagnosis method was proposed based on ACGAN and SDAE. The proposed method can reach a high accuracy even under small fault samples and have a good anti-noise capability compared with other bearing fault diagnosis methods. In [119], an ACGAN-based fault diagnosis method for rotary machines was proposed, whose performance was improved with the introduction of model agnostic meta learning. The experimental results demonstrated that the proposed method can significantly increase the fault diagnosis accuracy with limited training samples. In [120], an improved ACGAN was proposed to address the imbalance of dissolved gas analysis where the generator of ACGAN was constructed by using 1DCNN and long short-term memories. The experimental results indicated that the proposed method showed good performance for associated fault diagnosis with imbalanced data.

2.3.3 Other Variants of GAN

Deep convolutional generative adversarial networks (DCGANs) represent a significant advancement in generative models, integrating deep convolutional architectures into the GAN

framework to improve stability and representation learning. While primarily recognized for their ability to generate high-quality images, DCGANs are equally effective in processing one-dimensional signals, making them valuable for tasks such as time-series generation and fault diagnosis in mechanical systems, including bearings [121, 122]. By leveraging convolutional layers for feature extraction and adversarial training for data synthesis, DCGANs can model complex signal distributions, enhancing fault diagnosis accuracy and facilitating data augmentation for imbalanced datasets.

Cycle-consistent generative adversarial networks (CycleGANs) were introduced to address the challenge of unsupervised image-to-image translation, enabling the transformation between two domains without requiring paired data. Beyond their success in computer vision, CycleGANs have been effectively employed in the applications of bearing fault diagnosis [123, 124]. By leveraging cycle-consistency loss to preserve structural information during domain translation, CycleGANs can transform bearing signals from one operational condition to another, facilitating cross-domain diagnosis. This capability not only enhances the robustness of diagnostic models across varying conditions but also supports data augmentation by generating realistic fault signals for underrepresented scenarios.

Wasserstein generative adversarial networks (WGANs) were proposed to address the instability and mode collapse issues by introducing the Wasserstein distance as a more robust metric for measuring the divergence between the real and generated data distributions. In the context of bearing fault diagnosis, WGANs have been employed to synthesize realistic vibration signals, effectively augmenting datasets and addressing data imbalance challenges [125, 126]. By ensuring a stable training process and producing high-quality synthetic signals, WGANs enhance the performance of diagnostic models, enabling more accurate identification of fault types under varying operational conditions.

2.3.4 A Summary of GAN

A summary of GAN and the commonly used variants is concluded in Table 4.

Table 4. A summary of GAN and the commonly used variants.

Model	Advantages	Disadvantages
GAN	• Simplify implementation • Enable generation of realistic data samples.	• Exhibit training instability and mode collapse. • Produce variable-quality samples.
DCGAN	• Extract local features effectively. • Improve sample quality. • Stabilize training compared to standard GAN.	• Require adaptation for one-dimensional data. • Increase computational complexity.
ACGAN	• Provide additional supervisory information. • Ensure consistency between generated data and labels.	• Increase model complexity and training cost. • Classifier biases can affect the quality of generated samples.
CycleGAN	• Operate without paired datasets, enabling unsupervised learning. • Enables style transfer (e.g., normal-to-fault signal transformation).	• Require high computational resources and long training times. • May not always guarantee realistic cycle consistency.
WGAN	• Improve training stability and reduce mode collapse. • Provide meaningful loss measures to evaluate sample quality.	• Require careful tuning of hyperparameters. • Slow down training speed.

2.4 Digital Twin

DT technology represents a digital replica of a physical system that continuously updates with

real-time data from the physical counterpart. The primary advantage of DTs is their ability to simulate, monitor, and predict the behavior of physical systems, thus allowing for better decision-making and optimization [127]. Since the concept of DT was proposed in 2003, its content and connotation, structure, framework, implementation and so on have been extensively investigated and developed. Nowadays, DT encompasses not just the virtual representation of physical entities, it focuses on the exchange of data between physical entities and their virtual counterparts, as well as using the virtual entities to enhance the performance of the physical entities [128]. In recent years, DT technology has been adopted in various industries, including manufacturing, healthcare, and, increasingly, renewable energy, particularly in the wind energy sector [129]. An illustration of DT is shown in Figure 16. A twin indicates that the behavior or the state of the physical entity and virtual entity is equal to some degree. Once a change takes place in the physical entity, the change is delivered to the virtual entity by the signals or information. Then the virtual entity will adjust the parameters to reach equality. Similarly, once a change happened in the virtual entity, the change is delivered to the physical entity by the feedback information, therefore, the physical entity will adjust the parameters and get better performance.

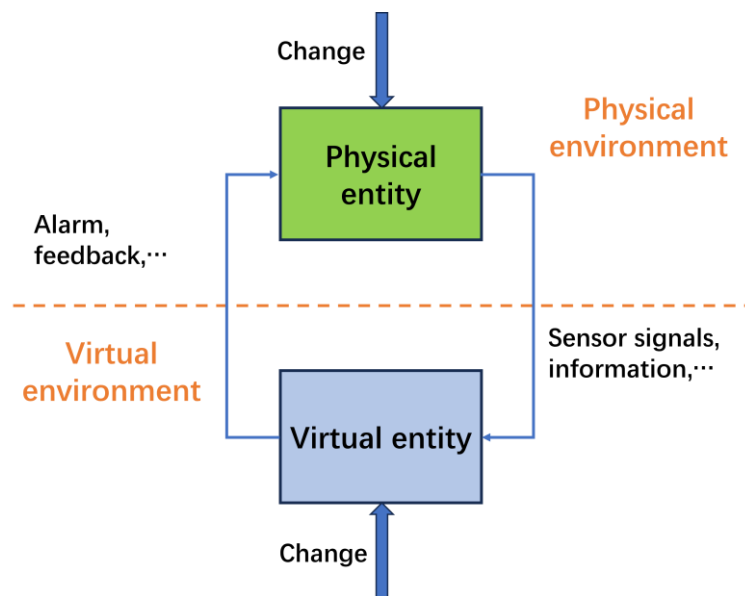


Figure 16. An illustration of DT.

Note the structure of DT is not fixed, it can be adjusted or changed to meet different specific application requirements. For example, [130] introduced a criterion to better describe DT after analyzing ten application areas of DT. It divided DT into five dimensions, namely, physical entity, virtual model, connection, data, and service, where the virtual model consists of four sub-models (i.e.,

geometric model, physical model, behavioral model, and rule model). Based on the criterion proposed in [130], Fei et al. analyzed 63 publications on manufacturing and found that only 29 of them included all the four sub-models [131].

The application of DT technology to WT systems offers several key advantages that significantly enhance the operational efficiency and reliability of WT systems including real-time monitoring, predictive maintenance, improving the reliability of WTs, extension of component service life, improvement of operational efficiency [132, 133]. First, DTs facilitate real-time monitoring by integrating sensor data from the turbine's various components, allowing for continuous assessment of performance and early detection of anomalies. This real-time insight enables timely intervention, reducing the risk of unexpected failures and improving overall WT safety. For example, in [134], the author proposed an assisted condition monitoring system based on DT for a 2 MW class double-fed induction generator-based WT which showed a good performance. In [135], researchers used a DT-based approach to simulate the real operation condition of WT drive train system and monitor the operation of WTs with the introduction of particle swarm optimization least squares SVM. The experimental results demonstrated that it can significantly enhance fault diagnosis accuracy and response speed, achieving a 99.1% success rate. Second, predictive maintenance is another major benefit enabled by DT technology. Through the analysis of historical and real-time operational data, DTs allow for the accurate forecasting of maintenance needs, thus optimizing maintenance schedules and minimizing downtime and associated costs. In [136], researchers had a detailed discussion about the distributed DT framework for predictive maintenance with a specific case study on WTs and demonstrated how the edge-fog-cloud paradigm can optimize DT implementation for predictive maintenance in WTs. Third, DTs play a crucial role in improving the reliability of WTs. By continuously updating virtual models with real-world data, they provide decision support that allows for more precise predictions of system behavior under various conditions, thereby reducing the likelihood of unplanned outages. In [137], researchers provided an overview of recent developments in the reliability analysis of offshore WT support structures, with particular emphasis on the incorporation of DT technology alongside finite element modeling. Thus, they proposed a novel DT framework that facilitates real-time monitoring and data integration, thereby enhancing the operational reliability and safety of offshore wind energy systems. Fourth, another notable advantage is the extension of component service life, achieved by using DTs to monitor fatigue and performance

degradation. This enables more efficient asset management, ensuring that WTs continue to operate effectively even as they near the end of their design life. Some researchers have discussed and proposed methods to predict the remaining useful life of WTs based on DT technology [138, 139] and they think DT technology can track and extend the lifespan of WT components. Finally, operational efficiency is significantly enhanced through the actionable insights provided by DTs, which optimize energy production, resource allocation, and overall wind farm performance.

A WT has many different sub-systems, for example, wind capture system, generator system, control system, and so on. The decrease in power generation due to the WT being in a sub-healthy state is just a superficial observation. Various faults of different subsystems can potentially contribute to this phenomenon. Therefore, despite the fact that one single DT can assess the state of the WT quite well, as will be shown in this study, it cannot provide more specific details of a fault, such the location or type of the fault. Due to these reasons, sub-DTs should be constructed based on different subsystems of the WT, and then these sub-DTs must be integrated into the overall DT constructed for the whole WT system. The relationship between WT and DT, together with their respective subsystems, is shown in Figure 17.

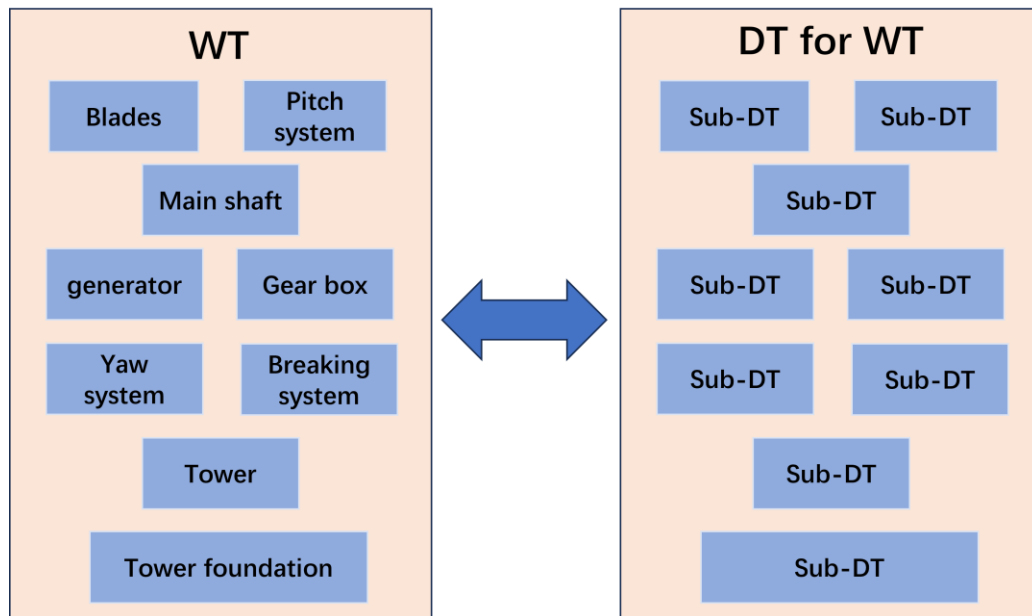


Figure 17. The relationship between WT and DT, together with their respective sub-systems.

DT technology offers a powerful framework for real-time monitoring, virtual simulation, and predictive analysis of complex physical assets such as WTs. Its capability to fuse multi-source data and physical models enables deeper insights into system behavior. However, most existing DT applications in WT diagnostics remain at the conceptual or simulation level, with limited integration

of intelligent fault detection and decision-making components. Furthermore, building high-fidelity DT models often requires extensive expert knowledge and significant computational resources. These limitations suggest a need for hybrid frameworks that combine DT with data-driven fault diagnosis methods to enable intelligent and scalable fault diagnosis.

2.5 Multi-Scale Feature Extraction

Multi-scale feature extraction is a powerful design strategy in neural networks that enables a model to capture a broader range of features by analyzing input data across multiple scales. This approach leverages parallel pathways or layers to process both fine-grained and high-level details, ensuring that diverse signal characteristics are preserved and comprehensively represented in the learned features. In the context of ACGAN, multi-scale feature extraction is implemented through a multi-pathway architecture in the generator and discriminator, where each pathway processes input data at a unique scale or resolution. Such a structural design enables the model to acquire a comprehensive feature representation, making it particularly effective for tasks like fault diagnosis.

Traditional DL-based methods and GAN-based methods often rely on a single pathway for feature extraction, which can limit their capacity to capture diverse characteristics in data, particularly for complex one-dimensional time series signals like vibration data [140]. By contrast, multi-scale approaches enable models to analyze information at various scales, capturing both fine-grained and broader patterns. This capability is critical for accurate classification in fault diagnosis applications, where signal variability may indicate subtle yet significant differences between healthy and fault states.

In the proposed multi-scale ACGAN model in Section 5.2, the generator and discriminator are restructured with multiple pathways that extract features across different scales, enabling a richer representation of the input data. This approach enhances model stability and improves classification performance by providing a more comprehensive view of the multi-dimensional characteristics of data. Multi-scale feature extraction allows ACGAN models to process vibration signals effectively, adapting to the unique demands of fault diagnosis in rotating machinery.

Multi-scale feature extraction has been proven essential in enhancing the effectiveness in fault diagnosis tasks. For instance, in [140], a novel WT bearing fault diagnosis method was proposed by combining the multi-scale feature extraction and CNN. By introducing multi-scale feature extraction, the proposed method can learn features at different scales which significantly improved the accuracy

of bearing fault diagnosis. In [141], a deep neural network was designed by employing multi-scale feature fusion thoughts for bearing fault diagnosis. The designed methods can obtain more abundant features from vibration signals and perform better. In [142], a new framework was developed by combining multi-scale down sampling and multi-scale feature extraction for rotating machinery fault diagnosis. The proposed fault diagnosis method can not only achieve higher accuracy but also have a good anti-noise ability. In [143], a novel 1DCNN-based bearing fault diagnosis method was proposed by integrating multi-scale feature extraction and attention mechanism. The proposed method can better extract the multi-scale deep information from the raw vibration signals, significantly increasing the diagnosis accuracy. In [144], a 2DCNN-based bearing fault diagnosis method was proposed by combining multi-scale feature extraction and attention mechanism. Different from that in [143], the framework of the designed model in [144] was based on the residual neural network, an improved 2DCNN. By introducing multi-scale blocks and attention mechanism blocks, the proposed method showed better reliability and effectiveness compared with other fault diagnosis methods. In [145], a multi-scale entropy features were extracted by combining stationary wavelet packet transform with both dispersion and permutation. The experimental results demonstrated that the proposed two fault diagnosis methods outperformed the stationary wavelet packet singular value entropy method. It has been observed that when applying multi-scale features to construct classification models, some features may be biased to large scale data, resulting in a negative effect on fault diagnosis performance. To address this problem, a novel multi-scale feature extraction method called Min-Max normalization was proposed accordingly [146].

Multi-scale feature extraction techniques enhance fault diagnosis performance by capturing information at various scales. This diversity improves the model's sensitivity to different fault types and degradation patterns. However, excessive or poorly selected multi-scale features may lead to redundancy and an increased computational burden. These issues highlight the importance of designing efficient multi-scale feature extraction methods, as further explored in this thesis.

2.6 Summary and Link to Research Gaps

This chapter reviewed key techniques for WT fault diagnosis, including traditional ML, advanced DL, DT technologies, and multi-scale feature extraction approaches. Each category contributes

valuable tools for capturing fault characteristics and improving fault diagnosis accuracy. However, several limitations remain unresolved.

Traditional ML methods rely heavily on handcrafted features and manual feature selection, which limits their adaptability to complex or non-stationary data commonly observed in WTs. DL methods, while capable of automated feature learning and high accuracy, often require large and labeled datasets. In addition, the accuracy of DL-based methods may decrease deeply when dealing with imbalanced data. DT frameworks offer a promising pathway for system-level modeling and simulation, but most applications remain conceptual or simulation-based, lacking real-time integration with intelligent diagnostic components. Meanwhile, multi-scale feature extraction enhances signal representation but can introduce computational burden and redundant information if not properly designed.

These limitations strongly align with the research gaps discussed in Chapter 1: 1) lack of a unified and intelligent framework for real-time state evaluation and maintenance scheduling (addressed in Chapter 3); 2) limited treatment and formalization of the sub-healthy state in WTs (addressed in Chapter 3); 3) heavy dependence on manual feature extraction in traditional ML-based diagnosis (addressed in Chapter 4); 4) diagnosis performance degradation under imbalanced data scenarios (addressed in Chapter 5). The proposed iSEMA system and fault diagnosis methods in this thesis aim to address these issues by combining DT modeling, traditional ML, advanced DL, and an improved multi-scale ACGAN.

Chapter 3

The Framework of a Novel Intelligent State Evaluation and Maintenance Arrangement System for WTs

3.1 Introduction

To reduce the influence of global warming and protect the environment of earth, people now are more inclined to choose clean energies instead of traditional fossil fuels. Wind energy as one of the clean energies now account for a quarter of all kinds of power generation [1]. Due to the high level of noise, WTs (WTs) are typically built in remote locations away from cities, such as mountains or offshore areas. Therefore, maintenance tasks are essential for the proper operations of WTs. According to [7], the cost of maintenance after failure is nearly three times the cost of regular maintenance. If early warnings are sent to the WT monitoring and management center, then the concerned team can address these warnings during regular maintenance work instead of waiting until the sub-healthy state develops or leads to more serious faults. This will significantly improve the efficiency of the WT and reduce any potential downtime.

DT was first proposed by Michael Grieves [127] in 2003 to better manage the lifecycle of physical products. After that, Michael Grieves and many other researchers have expanded the definition of DT which can be typically described as an integrated entity of a physical product, a virtual description of that product, and the data communication between them [128]. In recent years, DT has been applied in various fields and promotes positive developments in these areas. In [147], the applications of DT in energy storage were reviewed, demonstrating the positive developments with DT in this field and discussing the challenges of DT. In the field of manufacturing, DT can provide a real-time response of the manufacture system and increase flexibility [148]. In [149], the possibilities and challenges of DT for the construction of smart cities were comprehensively reviewed.

In addition to the aforementioned fields, DTs have also been employed in WTs to ensure their stable operation more effectively. A real-time planetary gear fault diagnosis method was proposed in [150] by combining the atom search optimization-SVM and DT which can significantly improve the operation of WTs. Kim et al. [151] utilize various environmental information to design a predictive model for offshore WT power generation based on DT. The proposed system enables an accurate representation of the offshore WT power generation and makes contributions to the safety of the power system.

To characterize the dynamic behaviors of WTs, mathematical equations or models based on first principles can be used to construct virtual representations of physical entities. For example, in [151] five equations are used to describe various behaviors of the WT system, while in [152] around ten equations are used to describe bearings. In both [9] and [10], the first-principles models are used to build a DT which can significantly facilitate the study of the original physical systems (e.g. to generate data via numerical simulations by considering a variety of environmental conditions). It may not be highly difficult to obtain mathematical models to describe the behaviors of WTs for some specific tasks, but for many real application scenarios where WTs are installed in harsh environments with severe climatic conditions, it may not be straightforward to build reliable mathematical models. As an attractive alternative, data-driven modelling techniques, such as system identification [153-155], reduced-order models [156], joint load-response estimation based on unified linear input and state modelling methods [157], and ML methods [158, 159], provide a powerful tool which can learn and extract some useful information from data without need much a priori knowledge of the original physical systems; the learned information or extracted features can usually be used to perform fault detection and diagnosis tasks.

In this chapter, the framework of iSEMA system for WTs was proposed by innovatively combining DL and DT technology, where LSTM is employed to construct a virtual entity of DT. The motivation for proposing the iSEMA system is to better monitor the status of WTs and to schedule the maintenance more effectively. Typically, wind farm operators perform maintenance on a WT until the systems enters a period of low wind speed. If slight or mild anomalies in power generation can be identified in advance and addressed earlier, then multiple manual shutdowns may be prevented and avoided, thereby minimizing WT downtime and maximizing economic returns. The iSEMA system proposed in this study can effectively identify the “sub-healthy” state of WTs: a state where a

WT's performance has already deteriorated but obvious faults have not yet manifested. The system records the monitoring process and alerts wind farm operators once any "sub-healthy" state occurs. Operators can then judiciously schedule maintenance activities accordingly.

This study focuses on WTs operating under both steady and turbulent wind conditions, covering a range of wind speeds from low to high, as typically observed in inland and offshore environments. The system is designed to account for varying wind patterns, including turbulence intensity and vertical wind shear, which are critical for accurate performance assessment.

3.2 The Proposed Framework of the iSEMA System

The iSEMA system operates by integrating multiple sources of data including actual measured power out and environmental signals such as wind speed, direction, air density, and so on. These signals are gathered in real-time and compared against the predicted power output generated by the LSTM model. The system's decision-making is based on deviations between predicted and actual power outputs, thereby diagnosing the healthy or sub-healthy state of the WT. For example, once the power output drops below a predefined threshold (for example, 90%), the system evaluates this deviation over a series of days to determine if the WT is entering a sub-healthy state.

To characterize the dynamic behaviors of WTs, mathematical equations or models based on first principles can be used to construct the virtual representation of physical entities. For example, in [151], five equations are used to describe various behaviors of the WT system, while in [152], around ten equations are used to describe bearings. It may not be difficult to obtain mathematical models to describe the behaviors of WTs for some specific tasks, but for many real application scenarios where WTs are installed in harsh environments with severe climatic conditions [160], it may not be straightforward to build reliable mathematical models. As an attractive alternative, data-based modelling techniques such as system identification [153-155] and ML models [158, 159] provide a powerful tool.

LSTM has been widely used in time series data prediction, natural language processing, text recognition, and so on [161] due to its advantages of handling time series data and addressing the exploding gradient problems. This work is concerned with the description of the dynamic behaviors of WTs using various environment signals and predicting the generated power of WTs in different

environments. In this chapter, the generated power refers to the power produced by WTs in operational scenarios under health conditions. Data used in this thesis, such as power output and environmental conditions, are time-series in nature, and LSTM networks have been shown to outperform traditional methods in capturing such temporal patterns [162]. While traditional regression models or other ML models could be used for power prediction, they lack the ability to capture the temporal dependencies present in WT data. LSTM, on the other hand, is specifically designed to model such time-series data, making it a more suitable choice for predicting both short-term fluctuations and long-term performance trends.

The structure of the proposed iSEMA system is shown in Figure 18, where historical data under various physical environmental conditions (e.g., wind speed, direction, air density, air pressure, and temperature) can be collected from the associated meteorological masts around WTs, then they are divided into two parts, training data, and test data. After that, the training data and test data are separately fed to the mathematical model, which is the virtual representation of WTs. In this work, the mathematical model is constructed by LSTM, because LSTM has the advantages of handling time series data and addressing the exploding gradient problems. In addition, the criterion for determining whether a WT is in a healthy or sub-healthy condition is the difference between the actual power generated by the WT and the predicted power over a long time. Therefore, the output of this mathematical model is a prediction of sequence data and LSTM is good at handling these sequence problems [161]. The function of this mathematical model is to describe the behavior and assess the states of the WT by predicting the generated power of the WT under different environments. More details of the proposed network model including the settings of the associated hyper-parameters (e.g., numbers of hidden units and the optimization method used) are shown in Section 3.3.

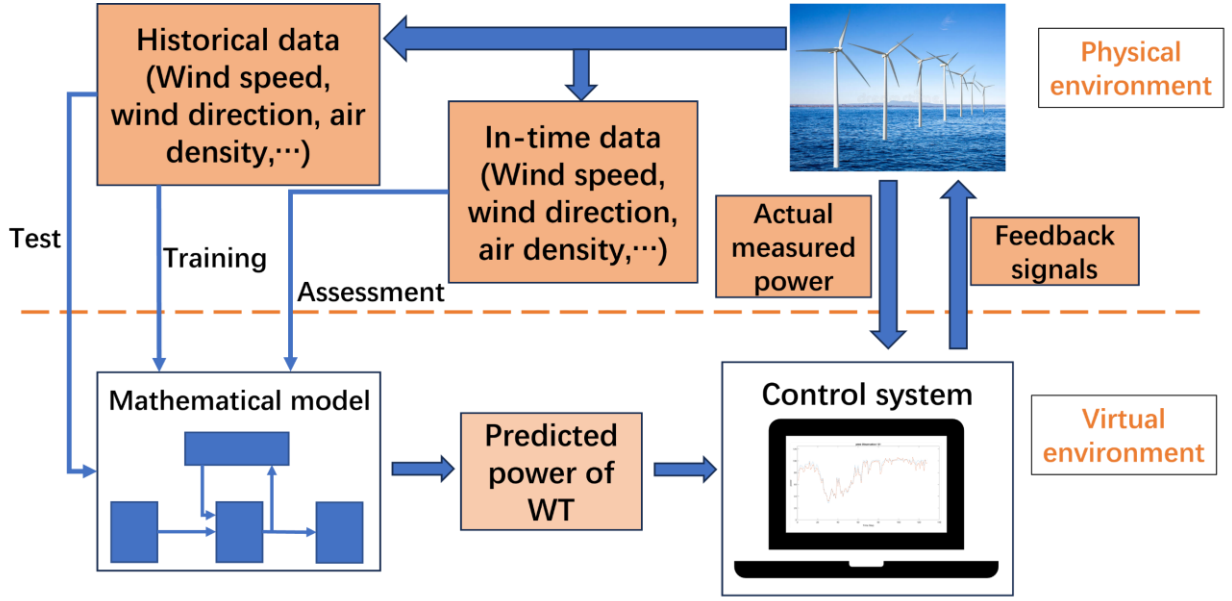


Figure 18. The structure of the proposed iSEMA system.

Once the mathematical model is constructed and well trained, collected online (in-time) data can then be fed into the model to assess the states of the WT. By comparing the predicted power and actual power, the proposed iSEMA system can distinguish the state of the WT which is either normal (healthy) or abnormal (sub-healthy). The system will send an alarm and information to operators if the WT is in an abnormal state. Follow-up maintenance can then be arranged based on the feedback signals.

To the best of our knowledge, there is no specific universal criterion to clearly define, characterize and discriminate the sub-healthy state of WTs. Typically, the sub-healthy state of a WT refers to an operational condition in which the WT is still functioning but has deviated from its optimal performance for a long period. In this state, WT exhibits signs of performance degradation or emerging anomalies that indicate potential risks, though these signs have not yet resulted in a complete failure or necessitated immediate shutdown. WT power production can be adversely affected by a variety of mechanical, electrical, and environmental factors. In particular, bearing degradation, gearbox faults, and generator inefficiencies can increase vibration, energy losses, and ultimately reduce power output. The proposed state evaluation model, which includes sub-healthy state recognition, is specifically designed to detect these early degradations before significant power loss occurs. Additionally, blade-related anomalies, such as aerodynamic imbalance or surface damage, can impair energy capture efficiency. External factors like low wind speed periods, turbine yaw

misalignment, or icing on blades can also cause temporary or long-term reductions in WT power production.

The introduction of the 'sub-healthy' state concept is pivotal for preventing severe failures in WTs. Traditionally, WTs are categorized as either 'healthy' or 'faulty', but this binary classification overlooks the early signs of performance degradation. By defining a sub-healthy state, the iSEMA system allows for more proactive maintenance scheduling, thereby reducing unplanned downtime and extending the lifespan of critical WT components. If early warnings can be sent to the WT monitoring and management center, who can then address these warnings during regular maintenance work instead of waiting until the sub-healthy state develops or leads to more serious faults, then it will significantly improve the efficiency of the WT and reduce any potential downtime. Although the term “sub-healthy” is not commonly used in industry, it shares conceptual similarities with “performance degradation” [6, 163], which is widely discussed in condition monitoring of WTs. As noticed in these works, the observed performance degradation in WTs may indicate various underlying issues across multiple components. Motivated by these works, this thesis introduces the concept of a sub-healthy state to emphasize the operational stage where performance degradation has begun but no observable fault has occurred.

The conceptual scope of sub-healthy state of WT includes:

1) Performance Deviation: The operational parameters of WT, such as power output (used in this thesis), rotational speed, vibration levels, or temperature, deviate from their designed optimal ranges, though the deviation is not severe enough to cause a breakdown.

2) Early Warning Signs: The sub-healthy state is characterized by the presence of early-stage anomalies that hint the onset of potential failures or accelerated wear, indicating that WT is at an increased risk of future malfunction if left unaddressed.

3) Need for Monitoring and Maintenance: While the WT can continue operating in a sub-healthy state, this condition needs close monitoring and preemptive maintenance to prevent further deterioration or failure.

4) Short-term Operability: Despite being in a sub-healthy condition, the WT remains operational without significant immediate impacts on its power generation capacity. However, continued operation in this state over time may accelerate wear and lead to major faults or system breakdowns.

The exclusions of sub-healthy state of WT are:

1) Critical failure or shutdown: Once the WT reaches a state where it requires immediate shutdown or repair, it is no longer in the sub-healthy state but in a failure or fault state.

2) Optimal health: In the healthy state, the WT operates within all specified design parameters, without any signs of performance degradation or potential faults.

In this thesis, the sub-healthy state of a WT is defined in two aspects, namely, the dimension of power and the dimension of time. To better define the sub-healthy state of the WT, it is necessary to introduce three relevant quantities first as follows:

CP: A criterion for determining whether the WT is in a sub-healthy state or a healthy state, which is typically represented as a certain percentage (CP).

Dmax: The maximum number of consecutive occurrences of sub-healthy days (D_{max}) in historical data determined by the system.

PSDP: The proportion of accumulated sub-healthy data points (PSDP) to the total number of data points in each day.

With the above quantities, the sub-healthy state of a WT can be described by the following three steps.

1. When the power of a data point at a specific moment is lower than a certain percentage (CP) of the predicted generated power, then that data point is seen as a sub-healthy data point.
2. If the proportion of accumulated sub-healthy data points to the total number of data points (PSDP) is larger than a specific value within one day, then that day is seen as a sub-healthy day.
3. If the situation continuously remains for $D_{max}+1$ days, the WT is seen as in a sub-healthy state.

In the iSEMA system, the time scale of one day (24 hours) is used as the measurement unit for the indicator D_{max} . Note that the total sub-healthy time determined by the system is fixed once the historical data were given. A smaller measurement unit (e.g. less than 24 hours), such as 12 hours, would not be a good choice because such a unit can increase the value of D_{max} .

For different types of WTs, the values of CP, D_{max} , and PSDP are usually different, however, typically, the value of CP should be specified or determined in advance based on a priori knowledge or empirical experience. In this work, CP is chosen to be 90%; such a threshold is adjustable for specific requirements and applications. This choice is informed by similar degradation prediction

studies in condition-based maintenance literature [164], where predicted performance dropping below 90% of the nominal value is often used as a signal of impending failure or functional decline. There is not a standard rule that can be used to choose the value of CP, in general the value should be less than 100% and it is not suggested to set CP to a value less than 70%. For a high CP value (e.g. exceeding 90%), the iSEMA system may not accurately determine the state of WTs. Furthermore, in practical scenarios, the actual power output of WTs may occasionally be slightly lower than the generated power, and usually this is not considered a faulty state. Therefore, a CP value that is very close to 100% is impractical. On the other hand, opting for a relatively low CP value, such as 50%, would imply that the system will not work until the actual power falls below 50% of the generated power. Again, this is an impractical or ineffective option. Once the value of CP is given, the proposed iSEMA system can automatically obtain the value of PSDP and Dmax and provide the information to operators in time. For convenience, it is necessary to introduce another indicator, called the ratio of adjusted power to the actual measured power (RoAP2MP) which is used to manually adjust the power of WTs during simulations to evaluate a sub-healthy state behavior. For convenience, the adjusted power refers to the power changed based on RoAP2MP and the measured power. Once the parameters and settings of the iSEMA system is determined, its performance can then be tested using the proposed models and methods. By changing the value of RoAP2MP, different states of WTs can be simulated. For example, by setting the value of RoAP2MP to 70%, simulations can be conducted to test if the iSEMA system can distinguish whether the WT is in a sub-healthy state. The algorithm of the diagnosis system is shown in Table 5.

Table 5. The algorithm of diagnosis system.

Step 1	Divide the historical data into three parts: training data, validation data, and test data, with a ratio of 6:2:2.
Step 2	Give the value of CP and set the value of RoAP2MP to CP+1%.
Step 3	Increase the value of PSDP from 50% to 100%, increasing by 5% each time to obtain the corresponding Dmax on training data.
Step 4	Obtain all the values of Dmax which are less than 10 along with their corresponding PSDP.
Step 5	Apply all the Dmax and corresponding PSDP on validation data and obtain the best pair of Dmax and PSDP which takes the shortest time to distinguish the sub-healthy state.
Step 6	Apply the best Dmax and PSDP on test data to test if it works.

Based on the above descriptions and discussion of defined quantities, it is clear that, in the iSEMA system, the sub-healthy state is determined using these quantities, such as CP, PSDP, and Dmax, which quantify deviations from optimal performance over time. CP evaluates the WT's power output efficiency, with values below 90% indicating that WT may be in a sub-healthy state. If the power output continues to be below CP and exceeds PSDP for a continuous day in statistics, then that day is considered a sub-healthy day. Note that, Dmax is the maximum number of consecutive occurrences of sub-healthy days in historical data determined by the system. Since LSTM's prediction of power output is not completely accurate, a normal day may be mistaken for a sub-healthy day. By introducing Dmax, it reduces the misjudgement of the WT's sub-healthy state caused by inaccurate LSTM predictions to a certain extent. Together, these metrics, CP, PSDP, and Dmax, provide a detailed understanding of the turbine's operational health and are critical for predicting maintenance needs. This allows the system to capture long-term trends in power output degradation, which might otherwise go unnoticed in traditional fault detection systems. By incorporating these parameters, the iSEMA system can provide early warnings of suboptimal performance.

3.3 Experiments and Results

3.3.1 Experiments and Data Description

The experimental design was guided by several key assumptions. First, we assumed that despite differences in environmental conditions, the fundamental operational patterns of WTs remain similar. This assumption allowed us to use the same model architecture for both inland and offshore WTs. Second, we hypothesized that the LSTM model, with its LSTM capabilities, would effectively capture the non-linear and time-varying dynamics of WT operations. The choice of Adam as the optimization algorithm was based on its superior experimental performance as confirmed by its lowest RMSE in comparison with other optimizers. Finally, considering that the WT lifecycle extends far beyond four years, it is assumed that the data of actual measured power output (inland and offshore power data) are the same as that of the generated power. That is to say, the WT is assumed to be in a healthy state for the first four years of its lifespan.

The selection of variables, including wind speed, wind direction, air density, turbulence intensity and so on, was based on their direct impact on the power generation and mechanical stress of WTs. These environmental factors are crucial for assessing WT performance and predicting potential faults. The LSTM model was chosen for its abilities to handle time-series data and capture long-term dependencies, which are essential for modelling WT behaviours. Furthermore, the 6:2:2 split of data into training, validation, and test sets was carefully selected to ensure sufficient data was available for training while retaining enough for unbiased validation and testing.

To validate the effectiveness and robustness of the iSEMA system, the experiments were designed to cover a variety of environmental conditions and WT types, including offshore and inland WTs. This approach ensures that the system can generalize well across different operational scenarios. The primary goal of the experiments was to assess the system's ability to diagnose the healthy and sub-healthy states of WTs under diverse conditions. The decision to include multiple types of WTs was motivated by the need to confirm that the iSEMA system could be generalized and applied to a broad range of real-world scenarios.

This section provides two case studies: one for offshore wind power data analysis and modelling, and another focusing on the data related to inland WTs. The work by Ding [165] provides the data

for both cases, that are measured and collected during the first four years of the turbine's operations. For the offshore WTs, data were recorded every 10 minutes from January 2007 to December 2010, including the ratio of the actual measured power to rated power and associated environment signals (wind speed, air density, humidity and turbulence intensity). For the inland WTs, data were recorded every 10 minutes from January 2008 to October 2011, including the ratio of actual measured power to the rated power and associated environment signals (wind speed, air density, turbulence intensity and vertical wind shear). Except for the wind speed and turbulence intensity were collected from the WTs, most of the environment signals were obtained from the around meteorological masts. For each case study, data were collected from the same type of WTs but with different locations. In this study, data from one WT are employed to construct the iSEMA system and data from other WTs (of the same type but installed at different location) are employed to evaluate the performance of the proposed system. To illustrate the daily variations of WTs, each sample comprises 144 data points.

The public data used in this section is collected in real practice WTs and associated meteorological masts. The coordinates of WTs and meteorological masts in the data are added with a constant by the real coordinates for the purpose of protecting their true geographic information. The authenticity ensures that the data is representative of the WT operational behaviors we aim to model, making them suitable for training and testing the iSEMA system. The data has been used and validated in studies because, on the webpage of this dataset, it can be found that the data has been viewed more than 2000 times and downloaded more than 500 times.

In this section, data from different sources were used without standardization. The decision not to standardize the data was made to preserve the original characteristics of the signals, which could be essential for capturing the nuanced behavior of WTs under varying environmental conditions. Although no explicit standardization was performed, the data used in this study share common characteristics. All power outputs were collected from operational WTs using similar sensor configurations and all environment information was collected from the closest meteorological masts, ensuring a level of consistency across different data sources. These shared attributes allowed us to analyze the data collectively without the need for further normalization, as the data exhibited common temporal and structural patterns.

3.3.2 Experiments on Offshore Data

3.3.2.1 Design of the Virtual Model

For the offshore case, there are two WTs, namely, WT3 and WT4. Data from WT3 are employed to construct the proposed iSEMA system and data from WT4 are employed to test the performance of the designed system. As described in Section 3.2, LSTM neural networks are used to build the virtual model, whose input signals are wind speed, wind direction, air density, humidity, and turbulence intensity. Wind speed is the primary driver of power output, while wind direction affects the alignment and efficiency of the WT blades. Air density plays a key role in determining the energy available in the wind, with variations affecting power generation efficiency. Humidity impacts the aerodynamic properties of the air and can contribute to material degradation, particularly in offshore environments. Finally, turbulence intensity significantly influences mechanical stress and fatigue, making it critical for predicting long-term turbine health. The unrolled structure of proposed LSTM networks is shown in Figure 19, where x_t , c_t , and h_t are the input data, memory data, and hidden data at time step t , separately. Because the length of each sample is 144, the total number of time steps is 144.

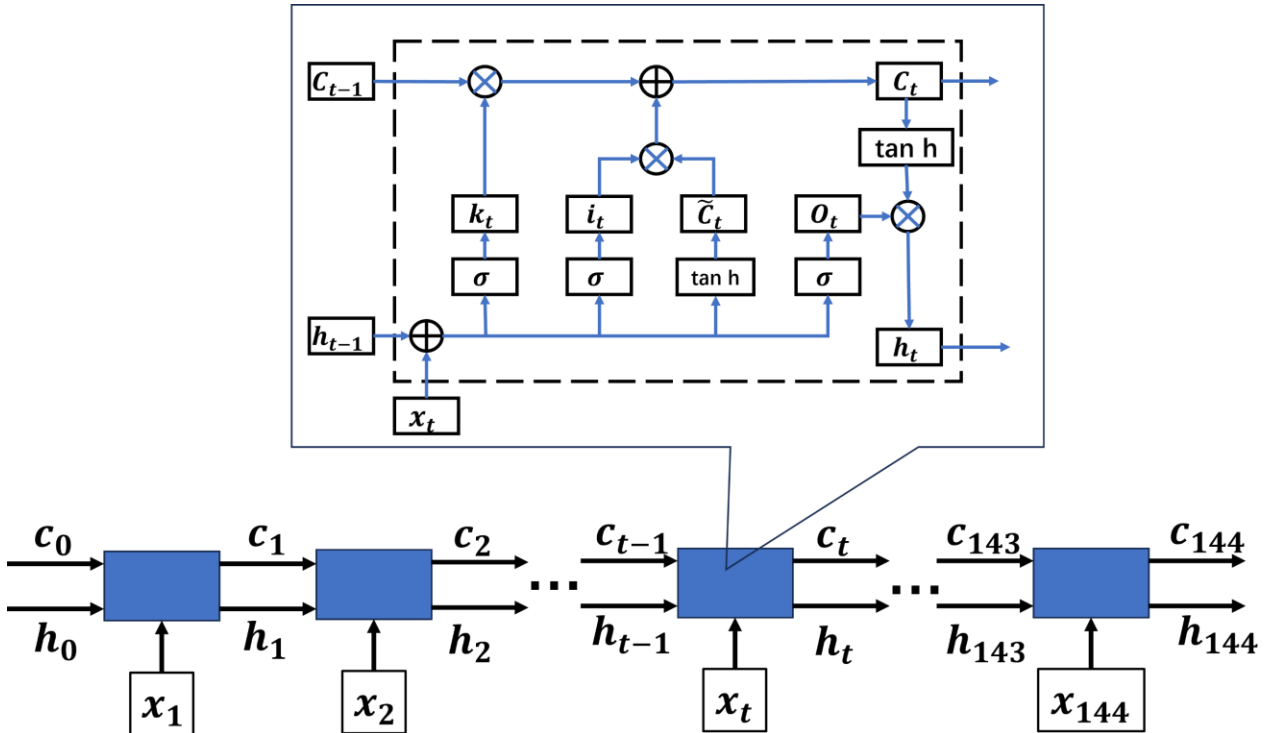


Figure 19. The unrolled structure of proposed LSTM networks.

To obtain an LSTM network model with best performance, a two-step scheme is performed to choose the most suitable optimization algorithm and determine the best number of hidden units of the LSTM networks. In step 1, the optimization method is chosen one by one from stochastic gradient descent with momentum (SGDM), root mean square propagation (RMSProp), and adaptive moment estimation (Adam), with other parameters being kept constant. SGDM was selected for initial model optimization due to its ability to accelerate convergence by incorporating a momentum term. This helps the model navigate areas of high curvature more efficiently and avoid local minima, making it suitable for training complex models like LSTM where gradient instability is a potential challenge. RMSProp was evaluated due to its ability to adaptively adjust the learning rate for each parameter, which is particularly useful in time-series data like WT performance. This method helps smooth out the learning process by controlling the update step for parameters that exhibit different levels of variability, thus enhancing model stability. Adam was ultimately chosen for its superior performance in balancing convergence speed and stability. By combining the advantages of SGDM (momentum) and RMSProp (adaptive learning rate), Adam consistently achieved lower root mean square error (RMSE) in our experiments, making it the optimal choice for fine-tuning the LSTM-based iSEMA system. Figure 20 demonstrates the RMSE of LSTM under different optimization methods for ten trials. It can be seen that the LSTM model trained with Adam performs best in terms of RMSE, with an average value of 18.71. While SGDM and RMSProp offered specific benefits in terms of momentum and adaptive learning rates, Adam provided the best overall performance, allowing for faster convergence and lower prediction error, thus becoming the preferred optimizer for this study.

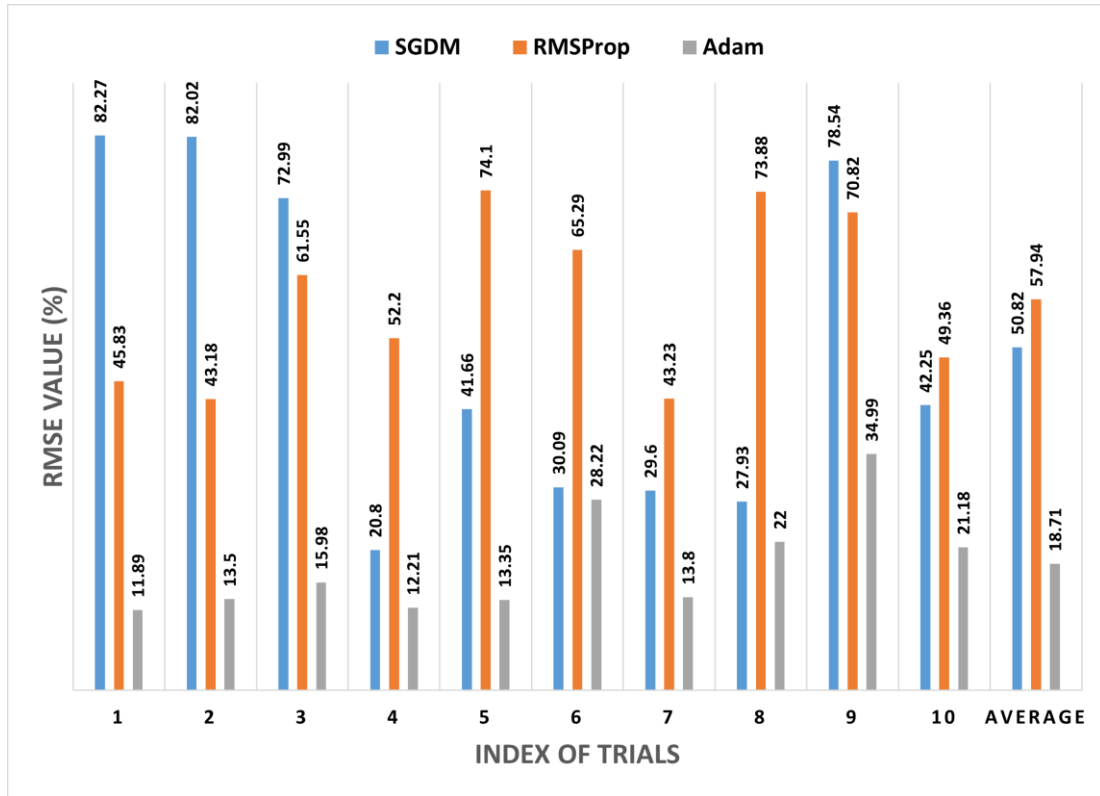


Figure 20. The RMSE of LSTM under different optimization methods.

Based on the experimental results of Step 1, the Adam optimization method is used for further model training. Thus, in Step 2, Adam is used to train different LSTM networks with different number of hidden units, ranging from 50 to 250 with an increase of 50 each time. For each of the cases (with different number of hidden units), a total of 10 experiments were carried out and the results are shown in Figure 21. From the experimental results, when the number of hidden units equals 150, the LSTM network shows the best performance in terms of the average RMSE value, which is 16.25. Note that among all the trained models, the LSTM network model with 150 hidden units shows the best performance, with an RMSE value of 8.85 in the last trial. Therefore, this LSTM model is selected as the virtual representation of WT3.

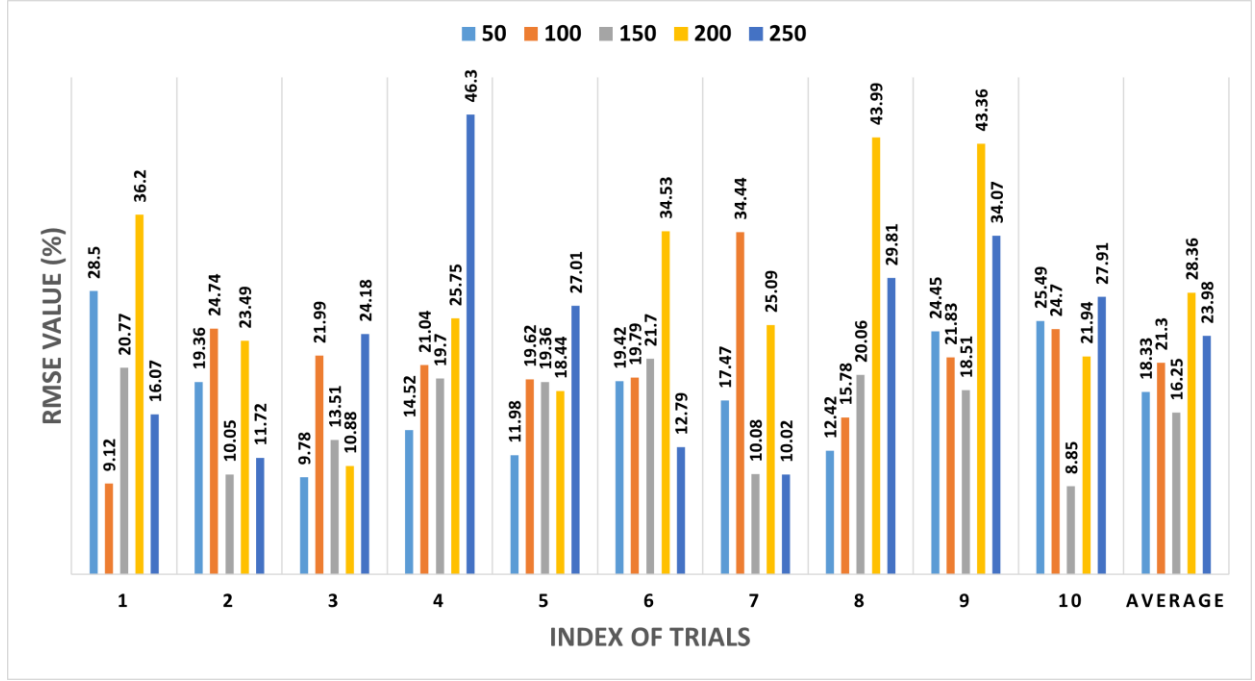


Figure 21. The RMSE of LSTM for different number of hidden units.

The distribution of the RMSE values of final selected LSTM model on the test data is shown in Figure 22, where it can be seen that the distribution of RMSE resembles a normal distribution. Figure 23 shows a comparison between the actual measured power and the corresponding model predictions on the offshore test dataset within one day. Here, ‘Power (%)’ denotes the ratio of the actual measured power of the WTs to the rated power or the predicted power by the LSTM model to the rated power. In most cases, the selected LSTM model has a good performance, but at certain times, the predictions are not accurate. For example, from Figure 23 it can be observed that the virtual model (LSTM networks) can predict the power generation very well but it can be noted that from the 50 to 100 data points, the actual values are higher than the predicted values.

To avoid the influence of imprecise predictions on WT state assessment and diagnosis, the two indicators, PSDP and Dmax, defined in Section 3.2, are introduced to the proposed iSEMA system. The local indicator PSDP, which is related to the accumulated sub-healthy data points, is insensitive to the bias in model prediction of the generated power in a short time as long as the prediction errors are small. For example, if PSDP is defined as 90% for the proposed system, then any proportion of misjudged data points less than this value can be avoided. By introducing the global indicator Dmax, the proposed system can address some extreme scenarios. For example, if the value of PSDP is small and the proportion of misjudged data points is larger than PSDP, then the WT state on that day would

be classified as in a sub-healthy state. Note that, the value of D_{max} is associated with the maximum number of consecutive occurrences of sub-healthy days in historical data. In this situation, D_{max} would become larger and allows the proposed iSEMA system to address such an extreme scenario.

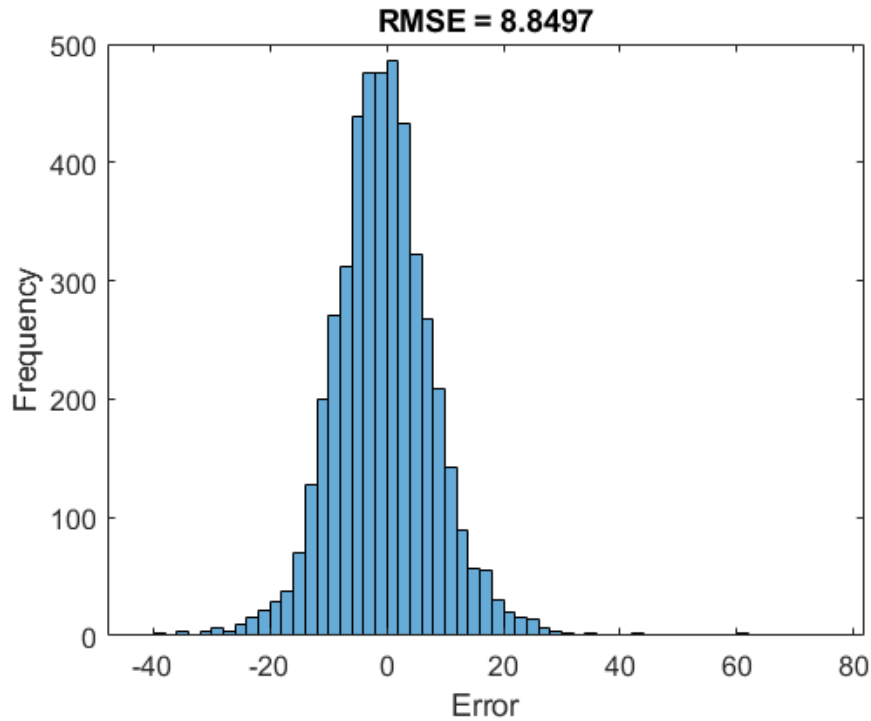


Figure 22. The RMSE distribution of LSTM in the test data.

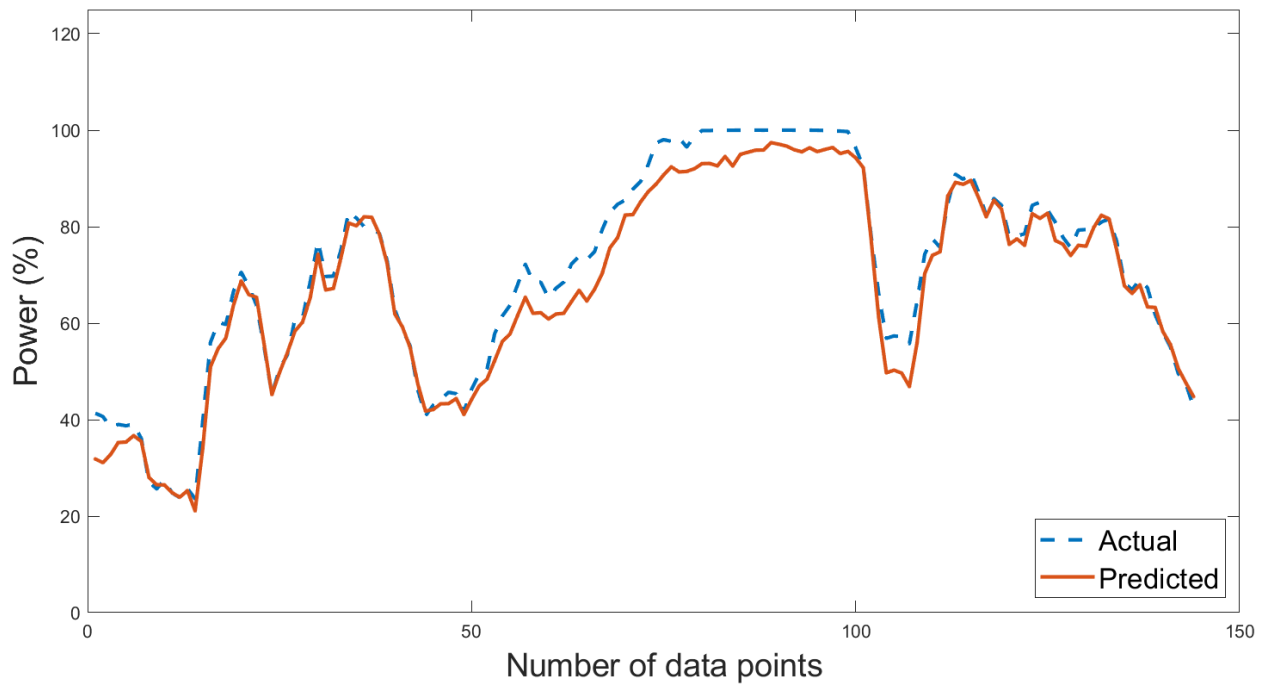


Figure 23. A comparison between the actual measured power and the predicted generated power in the offshore test dataset.

3.3.2.2 Selection of Indicators for the Diagnosis System

Once the LSTM model is well trained, the data are re-divided into three parts: training, validation, and test data with the ratio of 6:2:2 to determine the best value of indicators for the diagnosis system. The sample length is 144 and the number of samples for training, validation, and test data are 471, 157, and 157, separately. In this thesis, the certain percentage (CP) is set to 90% which means if the value of power data point at a specific moment is lower than 90% of the predicted value, then that data point is treated to be a sub-healthy point. As mentioned earlier, the global indicator Dmax is the maximum number of consecutive occurrences of sub-healthy days determined by the system, and PSDP is the local indicator related to the proportion of accumulated sub-healthy data points to the total number of data points in each day. Typically, the value of Dmax decreases with the increase of RoAP2MP when the value of PSDP is constant. To illustrate this, an experiment is carried out on the training data, with the following settings: the value of PSDP is set to be constant (50%); the value of RoAP2MP changes from 85% to 93% with an increase of 1% each time. The model was then simulated by considering different states of WT1. The resulting values of Dmax are shown in Table 6.

Table 6. The changes of Dmax with different RoAP2MP and PSDP being fixed at 50% for the offshore WT3 training data (unit: day).

RoAP2MP	85%	86%	87%	88%	89%	90%	91%	92%	93%
Dmax	89	84	46	28	27	17	17	17	17

The results in Table 6 demonstrate that the value of Dmax decreases with the increase of RoAP2MP when the value of PSDP is constant. In addition, when the value of RoAP2MP exceeds 90%, the value of Dmax keeps constant. If the parameter CP is set to be 90%, then when the proposed DT model iSEMA is used to simulate the WTs, all samples whose RoAP2MP is less than 90% will be treated as sub-healthy samples, resulting in a large value of Dmax. When WT is healthy, samples with RoAP2MP being larger than 90% will be seen as healthy samples. However, due to the errors in the LSTM model, some healthy samples may be misclassified as sub-healthy, resulting in a relatively small value of Dmax which could reduce the system's robustness. A smaller Dmax lowers the tolerance threshold for continuous sub-healthy days, meaning that even minor and temporary power

deviations, potentially due to prediction noise or environmental fluctuations, could cause the system to misjudge a healthy WT as sub-healthy. This leads to unnecessary early warnings, maintenance scheduling, or even interruptions in WT operation, which in turn affects operational efficiency and increases costs. Therefore, the selection of D_{max} should carefully balance sensitivity and specificity to avoid both delayed detection and false alarms.

To make the diagnosis system robust, reliable and effective, the following two considerations are put together in the iSEMA system design. Firstly, the diagnosis system should be able to effectively avoid misclassifying a healthy state as a sub-healthy state. Secondly, the system should be able to detect a sub-healthy state of WTs as quickly as possible. In doing so, when selecting the two parameters, D_{max} and PSDP, the resolution capability for situations where RoAP2MP is larger than CP (i.e., RoAP2MP is larger than 90%) is considered first. Furthermore, based on the results reported in Table 6, when RoAP2MP is larger than CP, D_{max} remains unchanged. Therefore, the selection of D_{max} and PSDP should be carried out under the premise of RoAP2MP being equal to $CP+1\%$, as shown in Table 5 (Step 2). In this way, most misdiagnoses of misclassifying a healthy state as a sub-healthy state can be avoided. For cases where RoAP2MP is larger than CP, but D_{max} decreases instead of remains unchanged, the premise of RoAP2MP being equal to $CP+1\%$ still works. This is because if a relatively larger D_{max} is chosen as the criterion for determining sub-healthy states, then the cases with a smaller D_{max} will be classified as healthy states correctly. For example, if RoAP2MP is set to 90% and assume D_{max} is 5, meaning that the WT is healthy, but it is on the verge of entering a sub-healthy state, then the maximum number of consecutive occurrences of sub-healthy days determined by the system is 5. Note that, if the situation of sub-healthy days persists for $D_{max}+1$ (i.e., 6) consecutive days, the WT is seen as in sub-healthy state. When RoAP2MP is larger than 90%, the value of D_{max} would be equal to or less than 5 which is smaller than 6. Therefore, a healthy WT state misjudged as sub-healthy is avoided. Based on the above discussions and the procedure given in Table 5, simulation experiments are carried out with the following settings: 1) the values of power are set to 91% of the actual measured power for all the samples on the training data, and 2) PSDP increases from 50% to 100%. The change of D_{max} is shown in Table 7.

Table 7. The change of Dmax with different PSDP and RoAP2MP being fixed at 91% for the offshore WT3

training data (unit: day).

PSDP	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
Dmax	17	17	17	15	15	12	12	11	7	4	2

The results in Table 7 demonstrate that when RoAP2MP is set to 91%, a larger PSDP value leads to a smaller Dmax. Note that if the situation of sub-healthy day continuously remains for Dmax+1 days as described in Section 3.2, then, the WT is seen as in a sub-healthy state. Clearly, the magnitude of Dmax directly influences the time required for determining the sub-healthy state of a WT. A prolonged evaluation period, in turn, may adversely impact the power generation or even WT shutdown, thereby incurring economic loss. The best values of Dmax and PSDP may be chosen from the last three pairs in Table 7, based which Dmax can be chosen as 7, 4 or 2. It is recommended that the value of Dmax should be smaller than 10. By choosing the last four pairs of Dmax and PSDP from Table 7, experiments are carried out on the validation data of WT3. When the value of PSDP changes from 85% to 100%, the condition cases are marked as A, B, C, and D, separately. The experimental results are recorded in Table 8. Note that a WT will usually enter into a sub-healthy state at a certain point after an extended period of normal operation, thereby resulting in a decrease in power generation. To sufficiently cover such realistic scenarios in the simulation experiments, the values of the WTs' power are adjusted based on RoAP2MP for part of the validation samples.

Table 8. Days required to determine sub-healthy state of WT3 for different RoAP2MP for the offshore validation

dataset (Unit: Day).

Cases	RoAP2MP									
	75%	80%	81%	82%	83%	84%	85%	89%	90%	95%
A	22	22	22	23	H	H	H	H	H	H
B	7	20	20	H	H	H	H	H	H	H
C	4	5	5	6	6	H	H	H	H	H
D	15	H	H	H	H	H	H	H	H	H

In Table 8, H means the state of WT3 is diagnosed as healthy by the designed iSEMA system. According to the results reported in Table 8, when the WT is healthy and the adjusted power maintains at or above 90% of actual measured power, the system can make the correct judgement for all the four condition cases. When RoAP2MP is within the range of 84% to 89%, it means the WT is in a sub-healthy state; the diagnosis system cannot correctly determine the WT state for any of the four condition cases. When the adjusted power decreases to 83% of actual measured power, only the condition Case C (Dmax is 4 and PSDP is 95%) can make the correct judgement. It takes six days from the point where the adjusted power consistently falls to or below 83% of actual measured power, until then the system confirms that the WT is in a sub-healthy state. Although the system performs best for condition case C among all the four condition cases, it still has an error when RoAP2MP is within the range of 84% to 89%. Compared with condition Case C, it usually requires a longer time for assessment or has a relatively larger error for other cases. Therefore, this work selects condition Case C, the most difficult tasks among the four, as a benchmark to show the performance of the proposed method. To test if condition Case C is suitable for WT3, experiments are conducted first on the test data and results are recorded in Table 9.

Table 9. Days required to distinguish sub-healthy state of WT3 for different RoAP2MP under condition Case C in the offshore test dataset.

RoAP2MP	75%	80%	81%	82%	83%	84%	85%	89%	90%	95%
Days	4	6	6	6	6	6	H	H	H	H

Experimental results in the test data demonstrate that when the WT is healthy and the adjusted power can be maintained at 90% of actual measured power or above, the diagnosis system can make the correct judgement. When the WT is in a sub-healthy state and the adjusted power decreases to 84% of actual measured power or lower, the system can determine that the WT is sub-healthy, but it takes six days to make a correct judgment, from the point where the adjusted power consistently falls to or below 84% of actual measured power until the system confirms that the WT is in a sub-healthy state.

To further validate the feasibility of the proposed system, experiments are conducted on the data collected from WT4. Note that WT4 is the same type as WT3, both are offshore WTs installed in

different locations. With the same model and experimental settings as for WT3, the results for WT4 are shown in Table 10.

Table 10. Days required to distinguish sub-healthy state of WT4 for different RoAP2MP under condition Case C.

RoAP2MP	75%	80%	81%	82%	85%	86%	87%	89%	90%	95%
Days	8	8	8	8	8	8	H	H	H	H

The results in Table 10 demonstrate that, the proposed system has a relatively worse performance for WT4 in comparison with that for WT3. This may be attributed to the operating and environment conditions. When WT4 is healthy, the system can make correct judgement. When the WT is sub-healthy and the adjusted power decreases to 86% of actual measured power, it takes eight days for the system to determine the sub-healthy state. The number of days required for determination is increased, but the error decreases.

After the evaluation of WTs, if the WT is in a sub-healthy state, the proposed iSEMA system can send an alert to associated staff (operators or workers). In addition, the proposed iSEMA system can provide an in-time comparison chart between the actual measured power and the predicted generated power as shown in Figure 24. By making use of the alerts and in-time comparison chart provided by the system, operators can proactively schedule maintenance work during periodic maintenance, and address potential issues and significantly improve the efficiency of WT.

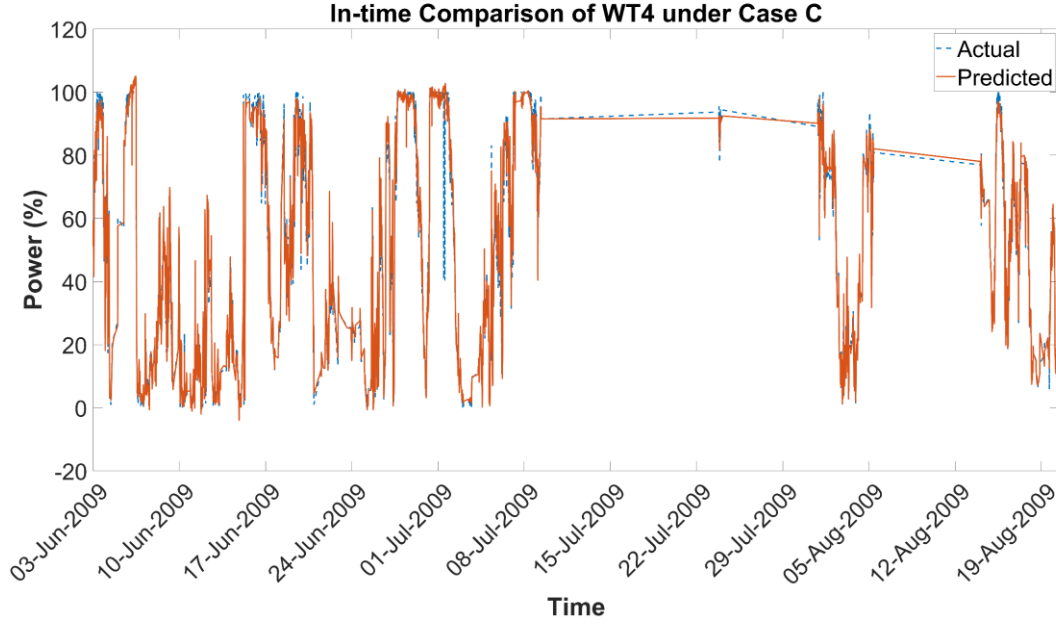


Figure 24. The in-time comparison chart produced by the iSEMA system of offshore WT4 under condition Case C.

3.3.3 Experiments on Inland Data

For inland data, there are also two WTs, namely WT1 and WT2. Data from WT1 is employed to construct the iSEMA system and data from WT2 is employed to test the feasibility of the proposed system. Wind speed, wind direction, air density, turbulence intensity, and vertical wind shear were collected as input to construct the LSTM model, whose structure is shown in Figure 19. The difference in variables between the inland and offshore data arises from the inherent environmental differences between the two locations. Inland WTs are more influenced by factors such as humidity and turbulence intensity due to complex terrain and atmospheric conditions. In contrast, offshore WTs experience more consistent wind patterns, making variables like turbulence intensity, and vertical wind shear more critical. These differences are naturally present in the raw data and reflect the distinct operational challenges faced by WTs in each environment. To construct the LSTM model, the priority is to define the parameters of the LSTM networks by applying the same method described in Section 3.3.2.1; it follows the following two-step scheme: 1) choosing the best optimization method from SGDM, RMSProp, and Adam, and 2) determining the best number of hidden units. Finally, the number of hidden units is set to 150 and the optimization method used is Adam. To demonstrate the

performance of the LSTM model, a comparison between the actual measured power from inland WT1 and the corresponding model predictions is shown in Figure 25.

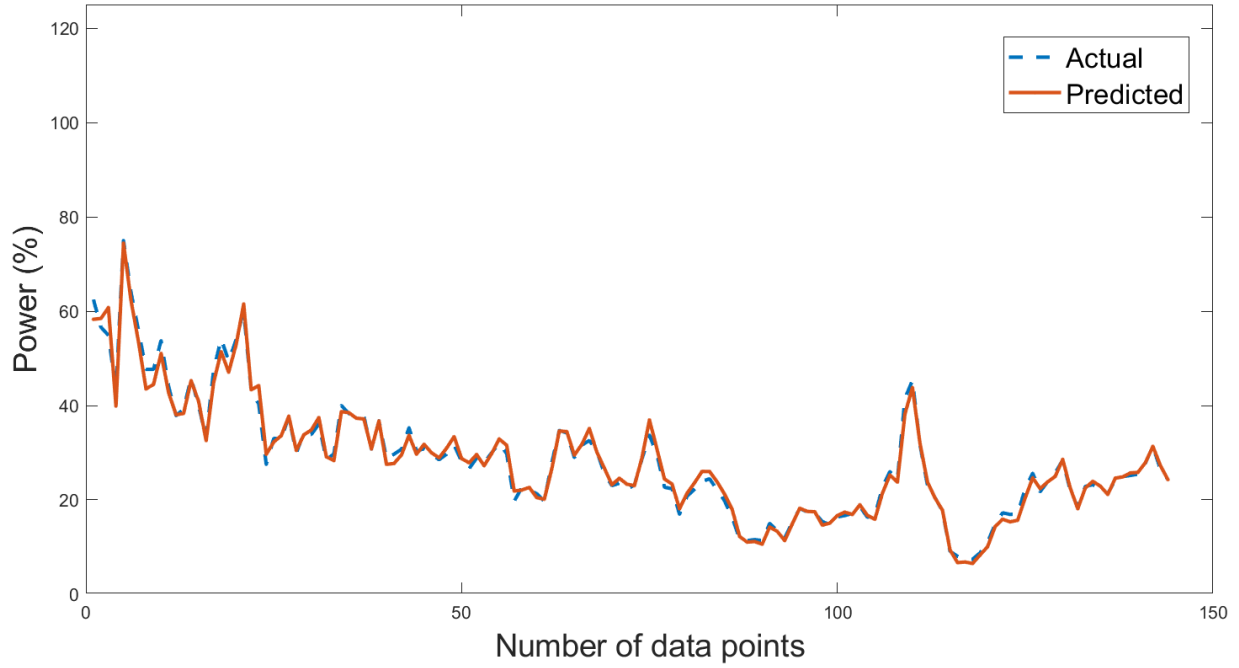


Figure 25. A comparison between the actual measured power and the predicted generated power for the inland test dataset.

Once the LSTM model is well trained, it can then be used for simulation studies. In this work, the experiments are designed as follows: 1) splitting the inland data into three parts: training data, validation data, and test data with the ratio of 6:2:2, 2) setting the value of CP to 90%, 3) obtaining the possible values of PSDP and Dmax in the training data, and 4) determining the best pair of PSDP and Dmax in the validation data. The experimental results on the training data are recorded in Table 11.

Table 11. The change of Dmax with different PSDP and RoAP2MP being fixed at 91% in the inland WT1 training data.

PSDP	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
Dmax	77	32	29	20	20	20	5	4	2	2	0

To enable the system to distinguish the state of WTs quickly, the value of Dmax should not be too large; it is suggested to be smaller than 10. Here, the last five pairs of indicators in Table 11 are

used as a reference to determine the value of PSDP. By using the last five values of PSDP in Table 11 and choosing 10 different values of RoAP2MP, numerical experiments were carried out on the data for WT1. Table 12 shows some results of the assessment of the state of WT1 under different conditions in the validation data.

Table 12. Days required to determine sub-healthy state of WT1 for different RoAP2MP in the inland validation dataset (Unit: Day).

PSDP	RoAP2MP									
	75%	80%	81%	84%	85%	86%	87%	89%	90%	95%
80%	5	5	5	7	7	7	H	H	H	H
85%	4	4	4	6	6	H	H	H	H	H
90%	2	4	4	4	4	6	H	H	H	H
95%	2	4	4	4	6	H	H	H	H	H
100%	1	3	5	5	H	H	H	H	H	H

When the adjusted power maintains at or above 90% of actual measured power, the system can make the correct judgement for all the five PSDP conditions (80% - 100%). Note that when RoAP2MP is below 90%, the WT state is defined as sub-healthy. From Table 12, it can be seen that when RoAP2MP is between 87% and 89%, meaning that the WT is in a sub-healthy state, the diagnostic system incorrectly classifies them as a health state. When the adjusted power decreases to 86% of actual measured power, the system makes the correct judgement for the first and third cases of five PSDP conditions (80% and 90%, respectively), but fails to make correct judgment for other cases. When the adjusted power drops to 84% of actual measured power, the system makes correct judgement for all the five cases. It can be noted from Table 12 that the diagnostic system performs relatively better for the third case (PSDP = 90%) than for the other cases. Therefore, the third case is used in further simulation study. Experiments are carried out on test data, and the results are recorded in Table 13.

Table 13. Days required to distinguish sub-healthy state of WT1 for different RoAP2MP under the third condition case on the test dataset.

RoAP2MP	75%	80%	85%	88%	89%	90%	91%	92%	93%	95%
Days	2	2	2	2	2	2	2	2	H	H

Experimental results on the test data demonstrate that when the WT is healthy and the adjusted power maintains at 93% of actual measured power or above, the system can make the correct judgement. However, when the WT is actually healthy and the adjusted power is within the range of 90% to 92% of actual measured power, the system classifies the WT state as in a sub-healthy. Such errors may be explained as follows: In Table 5, Step 3 is to find the best indicators (Dmax and PSDP) on the training data to avoid misclassifying a healthy state as a sub-healthy state, however, due to the uncertainty and changes in the internal and external working environment of the WTs, the distribution of test data may not be exactly the same as that of the training data, therefore the selected indicators based on the training data may not best fit the test data. For example, if the best value of Dmax should be 10, and PSDP should be 90%, but the selected Dmax is 6 and PSDP is 80% based on the training data, then errors may occur when the trained diagnostic system is applied to a test dataset. As a consequence, it may misclassify a healthy state as a sub-healthy state. When the adjusted power decreases to or below 89% of actual measured power, the system works better on the test data: it can correctly determine the WT state. It takes two days to make the correct judgment.

To further validate the feasibility of the proposed system, experiments are conducted on the data collected from WT2. Note that WT2 is the same type as WT1, both are inland WT but installed in different locations. With the same model and experimental settings as for WT1, the results on the whole data for WT2 are shown in Table 14.

Table 14. Days required to distinguish sub-healthy state for different RoAP2MP on the whole data of WT2 under the third condition case.

RoAP2MP	75%	80%	81%	82%	83%	84%	85%	89%	90%	95%
Days	2	2	2	6	9	9	H	H	H	H

The results demonstrate that the proposed system showed a relatively worse performance on

WT2 in comparison with that on WT1. The reason may be attributed to the change of operating conditions and work environment. When WT2 is healthy, the system can make correct judgement. When the WT is in sub-healthy and the adjusted power decreases to 84% of actual measured power, it takes nine days for the system to determine the sub-healthy state. The number of days required for determining the state is increased, and the range of error becomes larger. Overall, the proposed iSEMA system, trained with the data of WT1, works well for WT2. An advantage of the iSEMA system is that there is no need to re-train the LSTM model or re-select the values of indicators if the WTs are the same type.

3.4 Discussion about the Experimental Results of the iSEMA System

Using the proposed iSEMA system, it is possible to evaluate and predict the state of WTs by modelling and analyzing historical data of WTs. The selection of Dmax and PSDP is important for achieving quick and reliable diagnosis results as the two parameters have significant influence on the time required to distinguish the state of WTs. The proposed iSEMA system can distinguish the sub-healthy state of WTs quickly once the actual measured power of a WT decreases to a certain degree. Experimental results demonstrate that when the actual measured power maintained at or above 90% of the predicted generated power, the system can diagnose the healthy state correctly. However, when the actual measured power is set in the rough range from 85% to 89% of the predicted generated power, the diagnosis accuracy of iSEMA system is reduced. It was observed that, if the selected indicators are not good enough, misjudgment may occur as shown in Table 13.

Additionally, the model trained using the data of one WT can also be applied to other WTs of the same type, installed at different locations. This is important for improving the overall evaluation and maintenance efficiency for large scale wind farms, as models trained based on data of some WTs can be directly applied to other WTs without making major changes or adjustments to the pre-trained models DL models.

3.5 Conclusion about the Framework of the iSEMA System

In this chapter, the concept and definition of the sub-healthy state of WTs are introduced and expounded by introducing three parameters: CP, PSDP and Dmax. The framework of iSEMA system

designed based on DT technique is then proposed. The system can automatically evaluate the state of WTs and provide information that is useful for monitoring the state of WTs and scheduling maintenance more effectively. Thus, the efficiency of WTs can be highly improved.

The iSEMA system's performance was evaluated using data from both inland and offshore WTs. Each dataset was split using the 6:2:2 method. Metrics like RMSE and the number of days required to diagnose a sub-healthy state were used to measure model performance. The iSEMA system is highly adaptable and can be applied to various WT types and environmental conditions without retraining the model for each new site. Once the iSEMA system is trained based on data collected from one specific WT, it can be directly applied to other same type WTs in different locations. This flexibility comes from the fact that the LSTM-based model learns underlying patterns between WT behavior and environment information, which are transferrable between similar machines. The system is designed for ease of integration into existing maintenance workflows, enabling operators to preemptively address potential issues before they escalate into major faults.

A deficiency of the proposed iSEMA system is that when the measured power of WTs is within the rough range of 85% to 89% of predicted generated power, the model system may not always distinguish the state of WTs accurately. What's more, if the distribution of the test data is significantly different from the training data (which frequently occurs in real problem solving), then the selected indicators may not be good enough, as a consequence misjudgment of the WT state can occasionally occur. Therefore, it is essential to regularly adjust and update the associated parameters to maintain good performance of the system.

While the choice of CP is based on similar condition monitoring research and data analysis, there are some limitations. Firstly, this fixed threshold may not generalize well across all WT types or under highly variable environmental conditions. Secondly, the LSTM model's accuracy directly influences the reliability of the threshold; errors in forecasting can lead to either missed detections or premature warnings. In cases where a false alarm is triggered, the impact may include unnecessary maintenance interventions, increased operational cost, and potential loss of trust in the iSEMA system. However, such conservative threshold also helps avoid catastrophic failures. To mitigate false alarms, future deployments should consider integrating multi-source confirmation or apply adaptive thresholding mechanisms that adjust based on real-time trends and confidence intervals.

It is worth mentioning that the iSEMA system cannot provide specific information on the types

or locations of WT faults. In future, more experiments on sub-DT, involving key specific components of WT, will be carried out by incorporating different sub-DT systems into the iSEMA system, so as to enable the system to provide more specific details of faults and therefore significantly improve the accuracy of the iSEMA system.

Chapter 4

The Bearing Fault Diagnosis Method Based on Hybrid Framework

4.1 Introduction

Following the implementation of the iSEMA system in Chapter 3, which focuses on detecting the sub-healthy state of WTs, Chapter 4 shifts attention to a more specific fault category—bearing faults. While iSEMA provides a system-level anomaly detection based on power deviations, it does not explicitly identify fault types. Therefore, the bearing fault diagnosis model proposed in this chapter can be viewed as a sub-system of the iSEMA system that offers finer-grained insight into specific mechanical failures, thereby enhancing the reliability of the proposed iSEMA system.

Bearings are the key components of WT. Finding bearing faults in the early stage can not only keep WT running safely but also avoid economic loss. In the past, experienced experts could identify faults and find the locations by analyzing abnormal sounds or performance of bearings or employing statistical and signal processing methods, but it usually needed a huge amount of labor costs. However, nowadays it is unrealistic to ask the operators or maintainers to distinguish different types of faults manually. Conventional statistical methods and signal processing methods are easy to implement but when dealing with multi-classification tasks, the performance of these methods is usually not as good as ML or DL based methods. In addition, these conventional methods need manual interference, so it takes huge human labor costs to handle many different tasks [166, 167].

Intelligent fault diagnosis techniques, which can automatically build models using features extracted from raw data (recorded signals of the systems) to detect and diagnose faults with high accuracy, become necessary for real applications [60]. The fundamentals that enable intelligent fault diagnosis are ML techniques. In this thesis, ML is roughly divided into two groups: traditional or classical ML and modern deep neural networks or DL. The process of ML based fault diagnosis

comprises two parts: feature extraction and selection, and fault recognition and classification using ML methods [168]. Artificial feature extraction and selection usually need manual interference. For a given problem, different methods may produce different subsets of features. So, the choice of methods could have a great influence on the accuracy of the model. There are abundant ML models can be used for recognition and classification, for example, SVMs, decision trees, KNNs, ANNs. Although ML-based methods have many advantages compared with traditional manual interference methods, they still have some limitations. The manual interference in the process of feature extraction and selection is still needed, so these methods still rely on expert knowledge. What's more, modern industrial machines are becoming more and more complex, and more sensors are used to monitor the system or collect data. The performance of the traditional ML-based methods shows weakness when handling massive data.

DL based fault diagnosis can realize feature extraction automatically so there will be no requirement for expert knowledge. In addition, these methods have a good performance even handling big data in high dimensional space. Different signals have different characteristics and features. If we use expert knowledge to design algorithms to extract features and distinguish whether they are meaningful or meaningless for different tasks, it will be very complicated. Considering the powerful capability of handing massive data and good performance in feature extraction of DL, it is a good idea to use DL as a feature extraction tool. DL usually has an output defined by the SoftMax function which is used to calculate the probability of each class. However, probability only provides a measure of the likelihood of each class but cannot give a clear physical meaning or explanation in real practice. Therefore, it is worthy to carry out investigations on the proposed hybrid framework to obtain and enhance the interpretability of the network models.

4.2 The Proposed Hybrid Model

The framework of the proposed method can be divided into two parts: 1DCNN for feature extraction and RF for fault recognition and classification. The flow chart of the proposed framework is shown in Figure 26. Original vibration signals are sampled first, then, they are arranged as data matrix. To evaluate the performance of the proposed method, 10-fold cross-validation is used. Therefore, the data matrix is divided into 10 parts equally and the experiments are repeated 10 times.

In each trial, 9 parts are used for training and the rest is for a test. The training data and test data are sent to the 1DCNN model directly without any signal processing steps for feature extraction. After the operation of convolutional layers and pooling layers, feature vectors obtained from the FC layer of 1DCNN are used to train RF for fault recognition and classification. Finally, the model is estimated based on the test feature vectors.

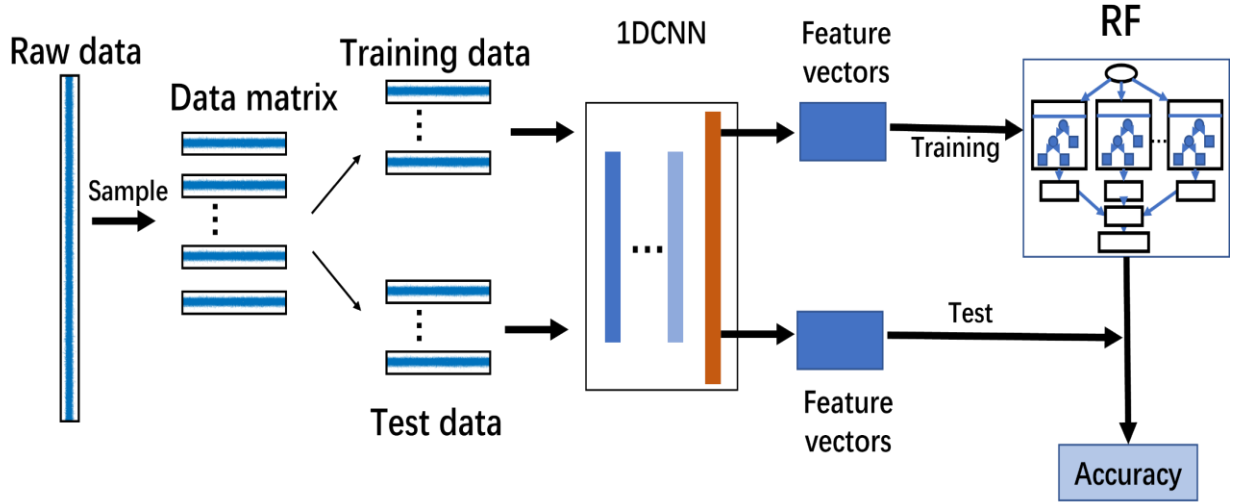


Figure 26. Flow chart of the proposed hybrid model

The hyperparameters in Table 15, including the number of filters and filter sizes, were initially selected based on the configuration presented in [62], which reported good performance in similar bearing fault diagnosis tasks. Generally, there is no criteria to select the hyperparameters of DL and the tuning usually takes huge amount of time, so it is a good idea to set these parameters based on this research. To verify their suitability, experiments were carried out with alternative settings, such as increasing the number of filters from 8 to 128 and varying the filter sizes from 2 to 32. However, these variations either resulted in negligible improvement or increased computational cost without significantly enhancing classification accuracy. In addition, when carrying out the experiment, it shows that if the number of convolutional layers is decreased to 3, it can still have a very good performance. When the number of convolutional layers comes to 2, it takes a long time to reach a high accuracy and when the number of convolutional layers comes to 1, the highest accuracy is only 60%. The final configuration shown in Table 15 was chosen as a balance between accuracy, training time, and model complexity. The parameters and structure of 1DCNN are shown in Table 15 and Figure 27, respectively, where ‘Conv’ represents the convolution layer, ‘Pool’ represents the pooling layer, and ‘FC’ represents the FC layer.

Table 15. Parameters of 1DCNN.

<i>Layer name</i>	<i>Filter size</i>	<i>Number of filters</i>	<i>Strides</i>
Conv1	16	128	1
Pool1	2	128	2
Conv2	8	64	1
Pool2	2	64	2
Conv3	4	32	1
Pool3	2	32	2
FC	10	1	

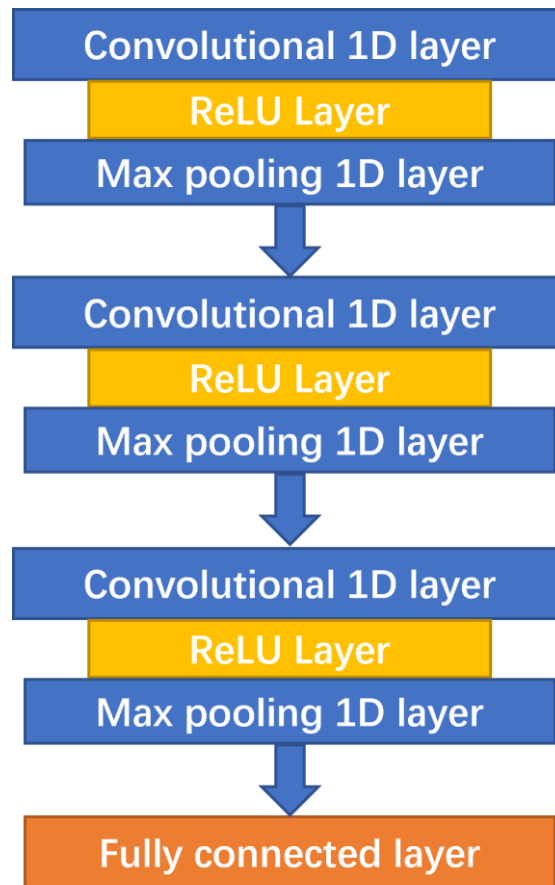


Figure 27. The structure of 1DCNN.

To prove the powerful feature extraction capability of 1DCNN, the t-Distributed Stochastic Neighbor Embedding [26] method is employed to visualize the outputs of each layer. In Figure 28, ten types of data are distributed in the space disorderly. They overlap and it is hard to distinguish. Then, they are sent to convolutional layers and pooling layers for feature extraction and the distribution is shown from Figure 29 to Figure 31. After the convolution and pooling operation, the disordered data points are gathered. Finally, in the FC layer, the extracted features from 10 types of

data can be distinguished clearly, as shown in Figure 32. Therefore, it is feasible to use 1DCNN as feature extraction method.

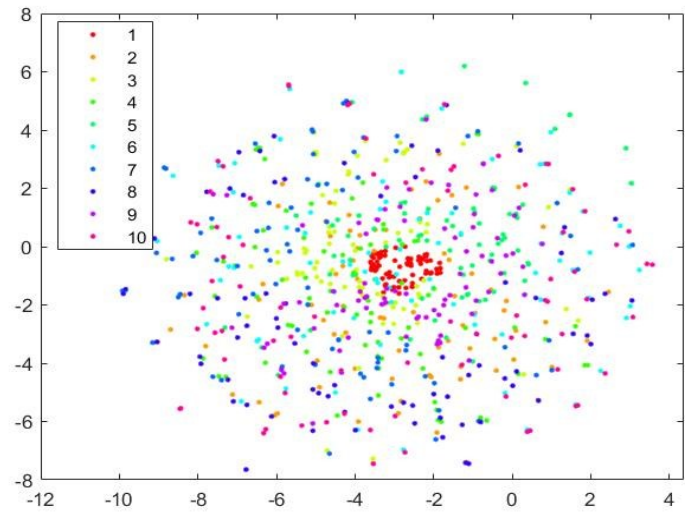


Figure 28. Distribution of raw data.

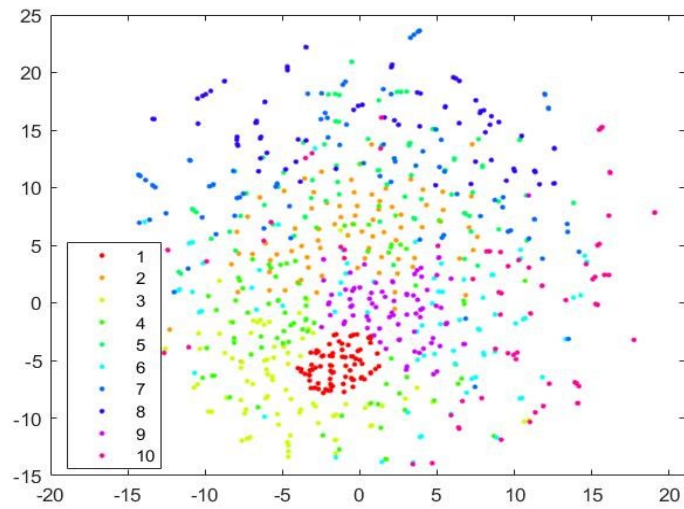


Figure 29. Extracted features distribution after Conv1.

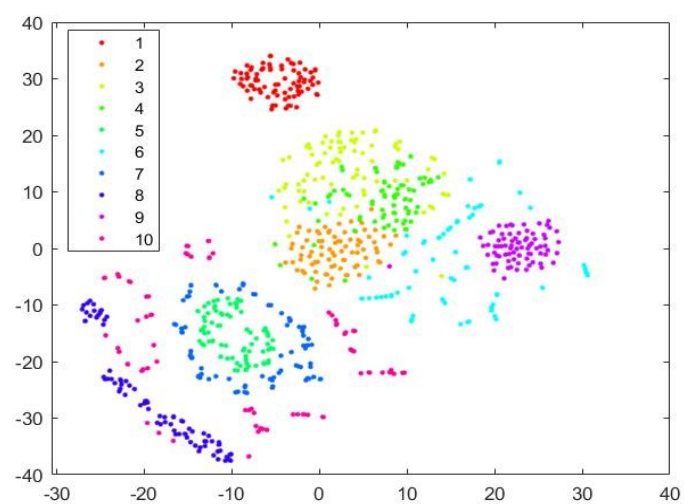


Figure 30. Extracted features distribution after Conv2.

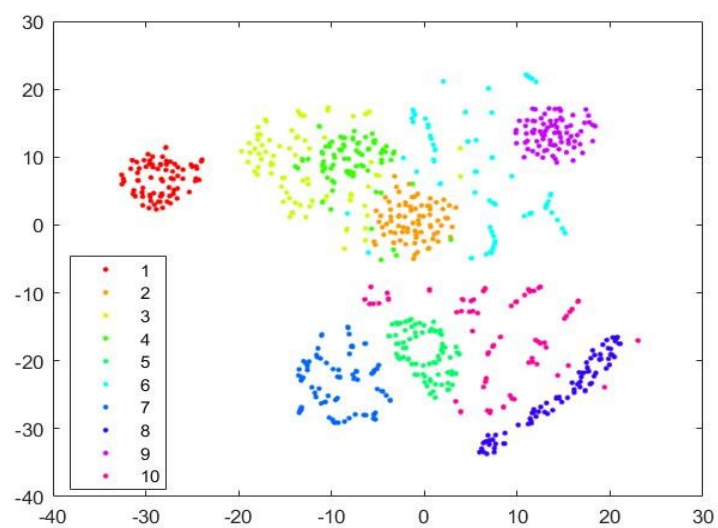


Figure 31. Extracted features distribution after Conv3.

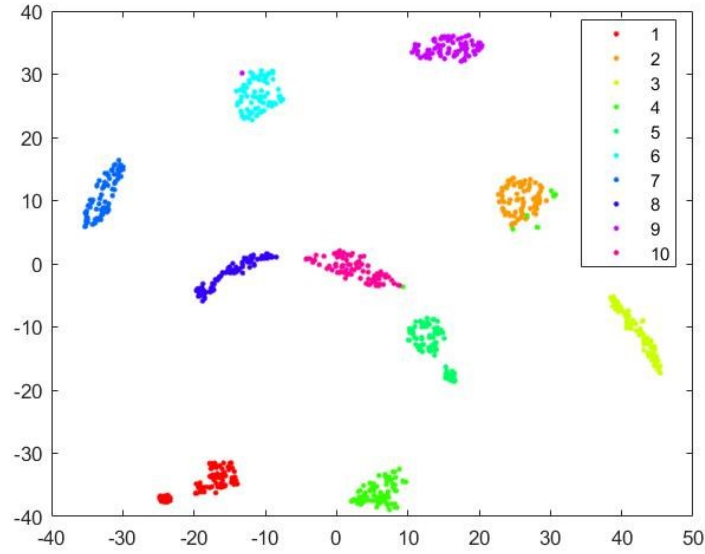


Figure 32. Extracted features distribution after the fully connected layer.

4.3 Experiments and Results

4.3.1 Experiments on the Case Western Reserve University Data

The raw data used in this case study is obtained from the Case Western Reserve University (CWRU) [169]. The data was collected by using a motor test rig equipped with a induction motor, a torque transducer, and accelerometers mounted on the motor housing. The data can be divided into four classes: 48kHz normal-baseline data, 48kHz drive-end fault data, 12 kHz drive-end fault data, and 12kHz fan-end fault data. In Section 4.3.1, the 48kHz normal-baseline data and the 12 kHz drive-end fault data were used. With different fault positions, there are three types of faults: ball fault, inner race fault, and outer race fault. Besides, there is one normal condition. All the data are collected under four different workloads, 0 HP, 1HP, 2HP, and 3HP with motor speeds varying between 1720 and 1797 RPM. In addition, for the fault data, there are three different fault diameters, 0.18mm, 0.36mm, and 0.54mm. In total, there are 10 different types of data, and the description of data can be seen in Table 16. Figure 33 presents four types of raw data in the CWRU dataset. To have better experimental results, 10-fold cross-validation was used. For each health condition, 800 samples are randomly selected, and each sample has 1024 points. To realize 10-fold cross-validation, each health condition

data is divided into 10 parts. In every trial, 9 parts are used as the training data and 1 part is used for the test.

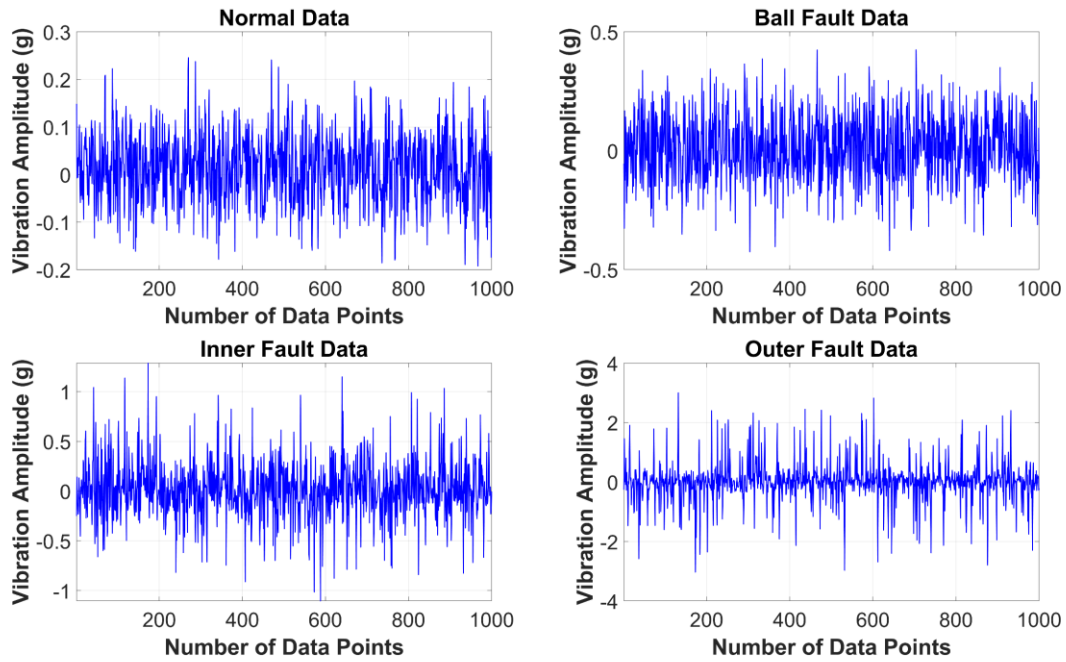


Figure 33. Presentation of raw data in the CWRU dataset.

Table 16. Data description of the CWRU bearing data.

<i>Health condition type</i>	<i>Label classes</i>	<i>Number of samples (training/test)</i>
Normal	1	720/80
BF-0.18	2	720/80
BF-0.36	3	720/80
BF-0.54	4	720/80
IF-0.18	5	720/80
IF-0.36	6	720/80
IF-0.54	7	720/80
OF-0.18	8	720/80
OF-0.36	9	720/80
OF-0.54	10	720/80

To construct traditional ML models for comparative experiments, the raw data must be processed by some signal processing methods to extract the features. Wavelet packet decomposition (WPD) is one of the commonly used feature extraction methods [170], it can decompose the raw data and provide the details and characteristics of raw data. In this chapter, the original signals are decomposed at layer 3 so 8 feature vectors are obtained. After that, principal component analysis (PCA) [45] is employed to reduce feature dimensions. The percentage of each component that describes the raw data are shown in Table 17. The criterion is to keep more than 95% of data information. From the table, it is obvious that component 1 to component 3 can describe more than 95% information. As a result, the 8-dimension feature vector is reduced to a 3-dimension feature vector. The model training process is shown in Figure 34.

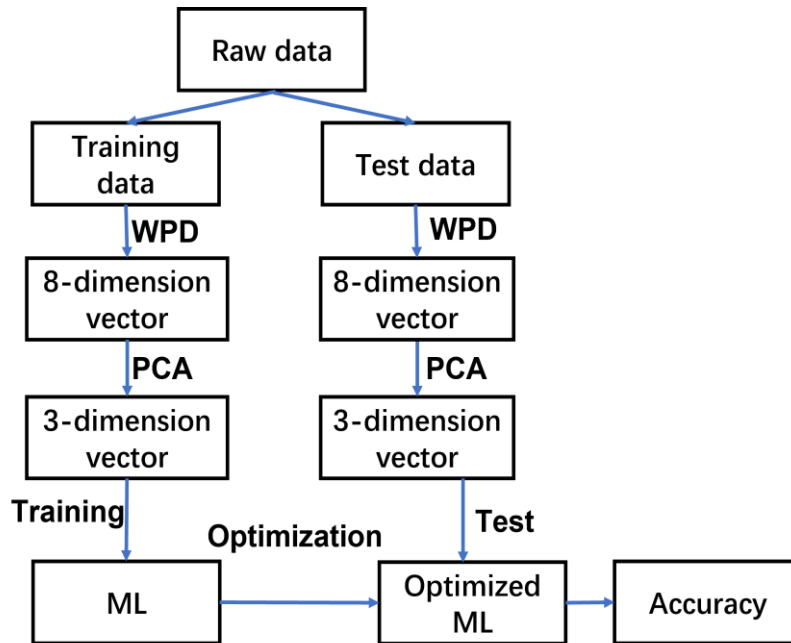


Figure 34. Flow chart of the traditional ML model.

Table 17. Explained percentage of 8 components on CWRU bearing data.

<i>Components</i>	<i>Percentage (%)</i>
Components1	77.14
Components2	15.17
Components3	4.24
Components4	2.67
Components5	0.65
Components6	0.11
Components7	0.01
Components8	2.65e-29

To make the experimental results convincing, 10-fold cross-validation is carried out. Each model is trained 10 times, and the results are recorded in Table 18 and Table 19. During the operation of traditional ML methods, some parameters should be confirmed. For example, the punishment factor C and the kernel function parameter g have a great influence on the performance of the SVM model. To illustrate this, experiments under the following two conditions were carried out using SVM: 1) using the default values $C = 1$ and $g = 0.1$; and 2) using the optimized values suggested by the GS algorithm [171]. For the operation of KNN, the number of neighbors K is searched from 2 to 50. During the training of RF, the number of trees is set to 500.

Table 18. Experimental results A on CWRU bearing data.

Number of trials	Accuracy (%)						
	<i>SVM</i>	<i>GS-SVM</i>	<i>KNN</i>	<i>RF</i>	<i>IDCNN-SVM</i>	<i>IDCNN-KNN</i>	<i>IDCNN-RF</i>
1	59.00	78.63	84.00	97.63	99.50	99.50	99.88
2	59.13	79.37	85.63	97.88	99.50	99.38	100.00
3	58.38	77.88	82.75	97.38	99.63	99.00	100.00
4	60.13	78.75	85.63	97.88	99.63	99.88	100.00
5	59.75	78.25	85.75	98.00	99.25	99.75	100.00
6	59.13	77.38	86.38	98.63	99.50	99.13	100.00
7	62.25	80.38	85.25	98.25	99.50	99.88	100.00
8	58.38	80.13	85.25	97.50	99.63	99.13	100.00
9	59.13	78.25	84.63	98.00	99.88	98.50	100.00
10	58.88	77.75	84.00	97.38	99.13	99.75	100.00
Average	59.42	78.68	84.93	97.86	99.52	99.39	99.99

It can be seen from Table 18 that RF performs best among all the traditional ML methods, but the accuracy is still lower than other hybrid methods. In addition, all hybrid methods are obviously significantly superior to the traditional ML based methods. Although the GS-SVM model significantly improves the accuracy of the conventional SVM model, running the model is very time consuming due to the optimization procedures. More importantly, the optimized GS-SVM model still has a large gap in performance compared with the hybrid model. Among all the hybrid methods, 1DCNN-RF performs best, and the average accuracy is 99.99%, even other hybrid methods perform quite good. In addition, RF has a clear structure and can address the overfitting problems, so the proposed 1DCNN-RF is superior to other methods.

Table 19. Experimental results B on CWRU bearing data.

Number of trials	Accuracy (%)		
	<i>GRU-SVM</i>	<i>GRU-KNN</i>	<i>GRU-RF</i>
1	100.00	100.00	100.00
2	100.00	100.00	100.00
3	100.00	100.00	100.00
4	100.00	100.00	100.00
5	100.00	100.00	100.00
6	100.00	100.00	100.00
7	100.00	100.00	100.00
8	100.00	100.00	100.00
9	100.00	100.00	100.00
10	100.00	100.00	100.00
Average	100.00	100.00	100.00

RNN can also be used for feature extraction, but the traditional RNN cannot keep the long-term information. While LSTM can process both the long-term and short-term information, it is quite time consuming to train a LSTM. In this experiment, GRU is used for feature extraction and experimental results are recorded in Table 19. As mentioned in Table 3, CNN may lose time dependencies of sequential data and this can also be proved from Table 18 and Table 19, where these three GRU based methods get 100 percent accuracy in all the 10 trials.

The average accuracy of KNN and RF are 84.93% and 97.86 which is quite good, but the average accuracy of SVM and GS-SVM are very low. Good performances of GS-SVM have been reported in

the literature (see e.g., [25]), unfortunately, the model does not show the expected high accuracy. This may be explained as follows. Firstly, the extracted features by WPD and PCA cannot describe the characteristics of the original signal very well. This is also the shortcoming of some fault diagnosis methods based on other traditional ML models. When dealing with massive and high-dimension data, it is hard to select suitable features to describe the raw data. Secondly, the SVM model was initially invented for binary classification problems. In this study, there are 10 types of data (1 health and 9 faults) so SVM may not be able to achieve a high accuracy for this problem. To increase the accuracy, a multi-classification structure may be considered, for example, one against all and one against one [29]. The accuracy of 1DCNN-SVM and GRU-SVM is higher than SVM and GS-SVM even though the parameters C and g are set to their default values. It confirms the powerful feature extraction ability of 1DCNN and GRU. It has been shown that 1DCNN and GRU can be used for bearings fault diagnosis [66, 172], but when using the DL methods for bearing fault diagnosis between different datasets, fine-tuning must be considered. By combining DL and ML, it can not only improve the performance of ML but also provide a more appropriate way to extract meaningful features. However, during the training of GRU, the gradient explosion problem happened frequently. One of the solutions is to decrease the learning rate but it takes more time to train the model. Considering the gradient explosion problem and low training speed, the proposed 1DCNN-RF is superior to the GRU-ML methods.

4.3.2 Experiments on the Jiang Nan University Data

The raw data used in this case study is obtained from the Jiang Nan University (JNU) [173], which is a publicly available benchmark dataset commonly used for bearing fault diagnosis research. The test platform consists of a centrifugal fan system driven by a Mitsubishi SB-JR induction motor. The motor is a 3.7 kW three-phase induction motor operating at 220 V with four poles and a rated speed of 1800 rpm. The rotor is supported by two rolling bearings. The vibration signals were collected using acceleration sensors mounted on the drive end of a rotating bearing test rig. The sampling frequency is 50 kHz. In total, there are four types of health condition: roller element fault (REF), out-race fault (OF), inner-race fault (IF) and normal (N). All the data are collected with three different rotating speeds: 600 rpm, 800 rpm and 1000 rpm. Details about this data can be seen in Table

20 and the presentation of raw data is shown in Figure 35. To have better experimental results, 10-fold cross-validation was used. For each health condition, 1200 samples are randomly selected, and each sample has 2400 points. To carry out 10-fold cross-validation, each health condition data is divided into 10 parts. In every trial, 9 parts are used as the training data and 1 part is used for the test. Although the 2400-point segment corresponds to only approximately half revolutions of the bearing at 600 rpm, it was sufficient for fault diagnosis in the experiment's setup. This is because localized bearing faults (e.g., inner race, outer race, or roller defects) typically generate impulsive vibration responses that can be captured even within partial revolutions, especially under high sampling rates. In the experiments, longer windows (e.g., 4000 and 6000 points) were also tested, but no significant improvement was found for classification accuracy. Thus, the 2400-point window was selected to strike a balance between performance and computational efficiency.

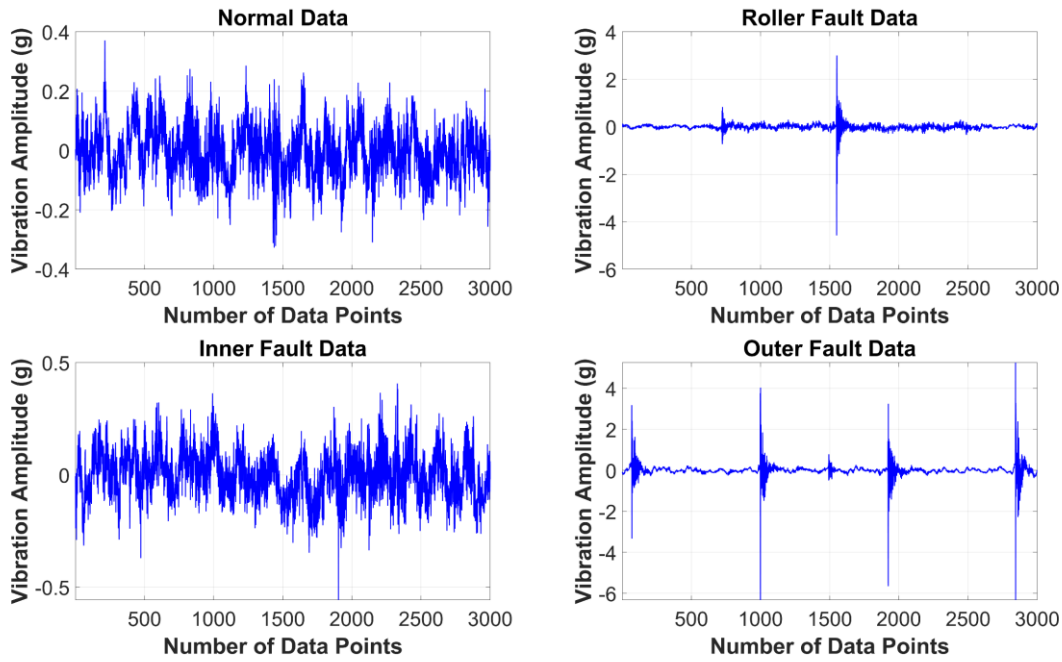


Figure 35. Presentation of raw data in the JNU dataset.

Table 20. Data description of the JNU bearing data

<i>Health condition type</i>	<i>Label classes</i>	<i>Number of samples (training/test)</i>
REF	1	1080/120
OF	2	1080/120
IF	3	1080/120
N	4	1080/120

The original signals are processed in the same way as in case study of the CWRU bearing data and the explained percentage of components is shown in Table 21. It should be noticed that, to describe 95% information, 4 components are needed. Therefore, the 8-dimension feature vector obtained by WPD is reduced to a 4-dimension feature vector.

Table 21. Explained percentage of 8 components on JNU bearing data.

<i>Components</i>	<i>Percentage (%)</i>
Components1	66.15
Components2	14.75
Components3	8.13
Components4	7.67
Components5	1.90
Components6	1.21
Components7	0.18
Components8	1.37e-29

To make the experimental results convincing, 10-fold cross-validation is also carried out and the results are recorded in Table 22 and Table 23. Similarly, 1DCNN and GRU are used for feature extraction first and the setting of other ML methods are the same as in case study of the CWRU bearing data. However, in this case study, there are only 4 types of health condition, so the default value of g in SVM is set to 0.25.

Table 22. Experimental results A on the JNU bearing data.

Number of trials	Accuracy (%)						
	<i>SVM</i>	<i>GS-SVM</i>	<i>KNN</i>	<i>RF</i>	<i>1DCNN-SVM</i>	<i>1DCNN-KNN</i>	<i>1DCNN-RF</i>
1	62.50	82.13	80.00	80.00	96.88	99.79	99.38
2	65.21	78.96	79.79	80.83	96.25	99.38	100.00
3	66.25	78.75	81.04	80.63	97.29	99.79	99.79
4	64.17	79.79	78.96	80.42	97.08	99.58	99.38
5	66.46	82.71	81.25	80.42	95.83	100.00	99.79
6	62.50	77.50	76.88	83.33	96.88	100.00	99.79
7	62.50	77.29	79.17	80.63	96.46	99.79	99.58
8	65.00	80.21	80.21	79.38	97.50	99.79	99.58
9	64.58	80.00	79.38	80.00	97.92	100.00	100.00
10	64.17	82.41	78.13	77.71	95.42	99.17	99.17
Average	64.33	79.98	79.48	80.34	96.75	99.73	99.67

Table 23. Experimental results B on the JNU bearing data.

Number of trials	Accuracy (%)		
	<i>GRU-SVM</i>	<i>GRU-KNN</i>	<i>GRU-RF</i>
1	99.79	100.00	100.00
2	99.79	100.00	99.79
3	100.00	99.79	99.79
4	99.17	99.38	98.96
5	99.79	100.00	100.00
6	99.58	99.58	99.79
7	99.58	99.79	99.58
8	99.38	100.00	99.79
9	100.00	100.00	100.00
10	99.38	99.79	99.38
Average	99.65	99.83	99.71

In this case, all the traditional ML methods have a bad performance where none of them can reach over 90% accuracy. The best one is RF, and the average accuracy is 80.34%. However, the accuracy of all hybrid models is over 96% which proves the feasibility of using DL as a feature extraction tool. Note that, in this case study, the gradient explosion problem also happened. Considering the training speed and high accuracy, the proposed 1DCNN-RF is superior to other methods.

4.4 Discussion about the Experimental Results of the Hybrid Model

The experimental results of the hybrid 1DCNN-RF model demonstrate its effectiveness in bearing fault classification across both the CWRU and JNU datasets. By integrating DL-based feature extraction and a traditional ML-based classifier, the model successfully leverages the strengths of both methods. The 1DCNN automatically extracts salient temporal features from raw vibration signals, reducing the dependency on manual signal processing. These features are then classified using a RF, which provides strong generalization capability and resistance to overfitting.

The hybrid model consistently achieves high classification accuracy across multiple fault types and signal conditions, which validates the robustness of the proposed approach under different data distributions. Compared with conventional ML models that rely on hand-crafted features, the hybrid structure offers a more adaptive and scalable solution. Furthermore, the model maintains interpretability and clear structure through the use of RF, making it more acceptable in industrial practice where transparent decision-making is valued. Overall, the hybrid framework presents a practical and high-performing diagnostic solution for real-world rotating machinery.

4.5 Conclusion about the Proposed Bearing Fault Diagnosis Method

In this chapter, the proposed hybrid framework is evaluated by two different datasets. The conclusion can be summarized that proposed hybrid framework can provide an efficient way of automatic feature extraction which can avoid manual interference during traditional ML models. In addition, the proposed framework can significantly improve performance of traditional ML models when dealing with multiple classification tasks. It should be noticed that the quantity of fault data is comparable to that of health data. However, in real-world scenarios, large machinery such as WTs are not permitted to operate in a faulty state for extended periods. This results in a significant disparity where fault data is far less abundant than health data. Utilizing such an imbalanced dataset for model construction and experimentation can substantially reduce the accuracy of the model. Solutions to address this issue will be presented in the next chapter.

These “100%” classification accuracies in the experiments demonstrates the strong classification capability of the proposed hybrid model on the used dataset. However, it is important to interpret these values within the context of the dataset and experimental setup. The used dataset was collected

under controlled laboratory conditions, with limited noise, fixed operating speeds, and clear fault classes. Therefore, the high accuracy primarily reflects the model's ability to classify well-structured and separable data. In real-world industrial applications, conditions are typically more variable and complex, involving multiple noise sources, mixed-mode faults, and changing loads and speeds, which can significantly affect model performance. For engineers, these results suggest that the proposed method is promising and performs well under ideal conditions. Nonetheless, further validation on diverse, noisy, and real-time data is required before the system can be confidently deployed in practice. These accuracy values should be regarded as indicative of the method's potential, not its guaranteed performance in all operational environments.

Chapter 5

A Novel Bearing Fault Diagnosis Method Based on Multi-Scale Auxiliary Classifier Generative Adversarial Network for Imbalanced Data

5.1 Introduction

Accurate bearing fault diagnosis is crucial as it directly impacts WT's efficiency and longevity. However, one of the primary challenges in this domain is the inherent imbalance in available data—healthy operating data is relatively easy to obtain, whereas fault data is often scarce. This imbalance can severely impair the performance of fault diagnosis models, leading to inaccurate predictions and unreliable maintenance strategies.

GANs and their variants, such as the ACGAN, have emerged as powerful tools to address data imbalance by generating synthetic data. However, existing GAN-based approaches often suffer from instability during training and may produce low-quality synthetic data, which limits their effectiveness in real-world applications [174].

In this thesis, we propose a Multi-Scale ACGAN designed specifically to improve the quality and diversity of synthetic fault data for bearing fault diagnosis. Our approach involves a comprehensive redesign of the generator and discriminator architectures, incorporating multi-scale features and optimized activation functions to enhance the generator's ability to produce realistic and varied fault data. The generated data is then used to train the classification model, 1DCNN, which is widely used in fault diagnosis tasks [175-177].

Extensive experiments validate the effectiveness of the proposed method, with results indicating that our Multi-Scale ACGAN not only outperforms traditional GAN and ACGAN-based methods but also provides a more robust and stable solution for bearing fault diagnosis. This work offers a significant step forward in addressing the challenges of data imbalance in fault diagnosis, providing a reliable and effective approach for improving the accuracy and reliability of diagnostic models in

real-world applications.

5.2 The Proposed ACGAN-Based Fault Diagnosis Method

5.2.1 A Brief Introduction to the Fault Diagnosis Tasks

The primary objective of this study is to address the challenge of bearing fault diagnosis in complex machinery systems, where data imbalance is a significant obstacle. Accurate fault diagnosis is crucial for the reliability and safety of machinery, as early detection of faults can prevent severe damage and reduce maintenance costs. However, real-world datasets often consist of a large volume of healthy data and a relatively small amount of fault data, which poses a challenge for conventional diagnostic models that require balanced data for effective learning.

The data used in this study consist of vibration signals representing both healthy and fault states under varying operational conditions. The data collected were initially balanced, containing adequate samples from both health and fault conditions across diverse fault types. To mimic real-world data imbalance scenarios and test the ability of the proposed methods to effectively solve fault diagnosis with imbalanced data, a portion of the fault data was deliberately removed, creating an imbalanced dataset. The resulting imbalanced data serves as the input for training the Multi-Scale ACGAN, which generates synthetic fault data. The combined real and synthetic data are then used for fault diagnosis.

The tasks to be addressed in this study are:

- **Data Augmentation:** Use Multi-Scale ACGAN to generate synthetic fault data, balancing the dataset to improve fault diagnosis accuracy.
- **Feature Extraction:** Implement a multi-scale feature extraction approach to capture diverse fault characteristics across different scales.
- **Model Evaluation:** Assess the effectiveness of the proposed method in accurately classifying faults and achieving robust performance under data imbalance.

5.2.2 Proposed Multi-Scale ACGAN

Gan [92] and ACGAN [112] were primarily proposed to better solve image classification tasks by extracting and making use of multi-scale features. In doing so, the structure of the generator and

discriminator were designed based on 2DCNN. In this chapter, however, the input data to be used are vibration signals which are presented in time series format; therefore, the structure of the existing generators and discriminators cannot be used directly but needs to be redesigned.

Inspired by the works reviewed in Section 2, this thesis proposes a simple but effective multi-scale feature extraction and fusion framework based on ACGAN, where the redesigned structure of the generator and discriminator is shown in Figure 36 and Figure 37, respectively. The FC layer is a basic and simple layer in neural networks. The generator and discriminator in this study are constructed using three parallel paths, namely, FC1 to FC2, FC3 to FC4, and FC5 to FC6. Simplified FC layers, due to their superior computational efficiency, are used to construct the parallel structure.

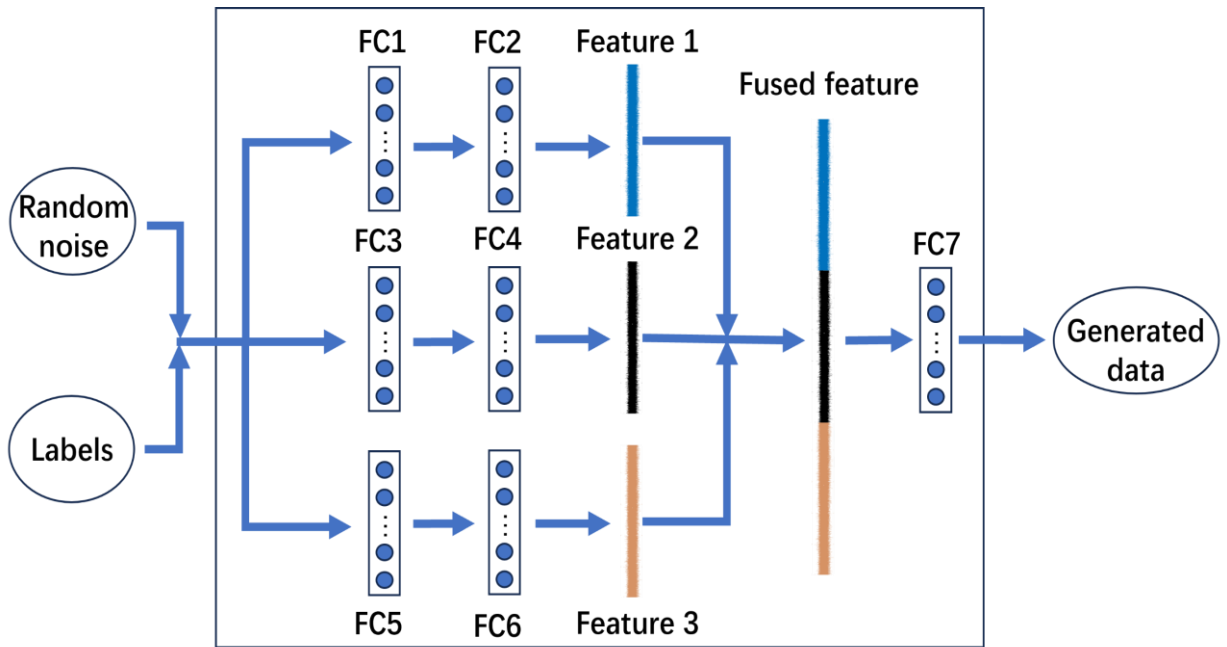


Figure 36. The redesigned structure of the generator.

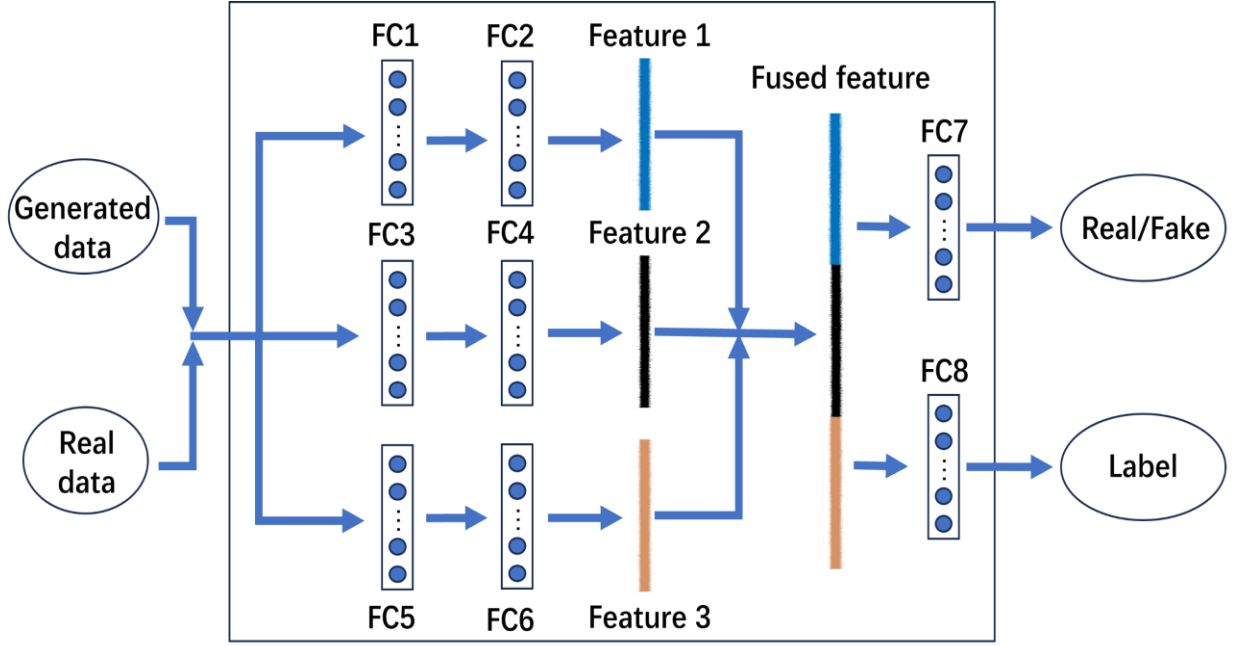


Figure 37. The redesigned structure of the discriminator.

Both the generator and discriminator share the multi-scale feature extraction approach, where the input data $\mathbf{I} \in \mathbb{R}^d$ (random noise in the generator, vibration signal in the discriminator) is processed through three parallel paths, each consisting of two FC layers. Therefore, these parallel structures can be described using the same core formula. The first FC layer transforms the input $\mathbf{I}$, and the process in the first FC layer of the path i can be described as:

$$\mathbf{F}_{i,1} = f(\mathbf{W}_{i,1}\mathbf{I} + \mathbf{b}_{i,1}), \quad \forall i \in \{1,2,3\} \quad (26)$$

where $\mathbf{W}_{i,1} \in \mathbb{R}^{m_1 \times d}$ and $\mathbf{b}_{i,1} \in \mathbb{R}^{m_1}$ are the weight matrix and bias vector for the first FC layer of path i , f is the activation function, and $\mathbf{F}_{i,1}$ is the extracted features by the first FC layer of path i . The second layer further processes the output of the first, and the process can be described as:

$$\mathbf{F}_{i,2} = f(\mathbf{W}_{i,2}\mathbf{F}_{i,1} + \mathbf{b}_{i,2}), \quad \forall i \in \{1,2,3\} \quad (27)$$

where $\mathbf{W}_{i,2} \in \mathbb{R}^{m_2 \times m_1}$ and $\mathbf{b}_{i,2} \in \mathbb{R}^{m_2}$ are the weight matrix and bias vector for the second FC layer of path i , f is the activation function, and $\mathbf{F}_{i,2}$ is the extracted features by the second FC layer of path i . These features from different paths are then concatenated to form a combined feature vector:

$$\mathbf{F}_{concat} = [\mathbf{F}_{1,2}, \mathbf{F}_{2,2}, \mathbf{F}_{3,2}] \quad (28)$$

This concatenated vector is then processed for different tasks. In the generator, the concatenated vector $\mathbf{F}_{concat}$ is processed by a final FC layer to generate the vibration signal output $G(\mathbf{F})$:

$$G(\mathbf{F}) = f(\mathbf{W}_g \mathbf{F}_{concat} + \mathbf{b}_g) \quad (29)$$

where $\mathbf{W}_g \in \mathbb{R}^{d_{out} \times 3m_2}$ and $\mathbf{b}_g \in \mathbb{R}^{d_{out}}$ are the weight matrix and bias vector for the final FC layer of generator, and f is the activation function.

In the discriminator, $\mathbf{F}_{concat}$ is processed by two FC layers. One is for distinguishing the vibration signal as real or fake, and the other one is for classifying the label information of the vibration signal. The processes can be described by:

$$D(\mathbf{F}) = f(\mathbf{W}_{dis} \mathbf{F}_{concat} + b_{dis}) \quad (30)$$

$$C(\mathbf{F}) = f(\mathbf{W}_{cla} \mathbf{F}_{concat} + \mathbf{b}_{cla}) \quad (31)$$

where $\mathbf{W}_{dis} \in \mathbb{R}^{1 \times 3m_2}$ and b_{dis} are the weight vector and bias of FC7 in discriminator; $\mathbf{W}_{cla} \in \mathbb{R}^{N \times 3m_2}$ and $\mathbf{b}_{cla} \in \mathbb{R}^N$ are the weight matrix and bias vector of FC8 in discriminator, N is the number of fault types; f is the activation function; $D(\mathbf{F})$ is the probability that the input $\mathbf{F}$ is real or fake and $C(\mathbf{F})$ is the result for the auxiliary classification task.

By incorporating the parallel FC layers at multiple scales, the model can capture a broader range of features across different dimensions, allowing for more comprehensive and detailed feature representation. This multi-scale approach enables both the generator and discriminator to extract and process diverse aspects of the data, which is particularly beneficial when dealing with complex signal patterns. In contrast, traditional GANs and ACGANs typically rely on a single pathway for feature extraction, limiting their capacity to fully capture the multi-dimensional characteristics inherent in the data. Additionally, the enhanced feature diversity enables better auxiliary classification performance, as the model can more effectively learn the nuanced distinctions between various signal classes.

To perform the multiple-scale feature extraction process, the FC layers are designed with different hidden units. In this thesis, the generator and discriminator are specially designed for processing vibration signals, so the final layers of them are different from the existing networks such as those proposed in [112]. Details of the generator and discriminator are shown in Table 24 and Table 25, respectively, where ‘No. of classes’ means the number of fault types. A sigmoid function is used in the discriminator to classify whether the data is real or fake. It is a binary classification problem, so Sigmoid is a good choice. In the discriminator, SoftMax is used to distinguish the types of faults from both real and fake data. The Tanh activation function is used in the generator to produce data

with values constrained between -1 and 1. This range is particularly effective for stabilizing GAN training, as it helps prevent extreme values in generated outputs and ensures smoother gradients, thus enhancing the generator’s ability to produce realistic synthetic data. More details about the influence of different activation functions on the proposed method can be found in the following section.

Table 24. Details of generator setting.

Layer	No. of hidden units	Activation function
FC1	15	ReLU
FC2	30	ReLU
FC3	10	ReLU
FC4	20	ReLU
FC5	5	ReLU
FC6	10	ReLU
FC7	No. of classes	Tanh

Table 25. Details of discriminator setting.

Layer	No. of hidden units	Activation function
FC1	25	ReLU
FC2	50	ReLU
FC3	15	ReLU
FC4	30	ReLU
FC5	10	ReLU
FC6	20	ReLU
FC7	1	Sigmoid
FC8	No. of classes	SoftMax

In the proposed multi-scale ACGAN framework, different numbers of hidden units of the generator and discriminator were deliberately assigned to each parallel path within the FC layers. This design choice is motivated by the goal of extracting features at multiple scales. Larger hidden layers tend to capture higher-level abstract patterns, whereas smaller ones focus more on localized or

fine-grained features. By configuring each path to have a distinct number of hidden units, the model can simultaneously learn complementary features across multiple resolutions. This architectural diversity enhances the representational capacity of the generator and discriminator, particularly in handling imbalanced fault data where subtle fault characteristics must be amplified.

The flowchart of the proposed ACGAN-based fault diagnosis method is shown in Figure 38 which starts from the bottom left corner. Initially, the raw data, consisting of healthy data (H) and multiple types of fault data (F1 to FN), is divided into two subsets: Data 1 (10%) and Data 2 (90%). Data 1 is reserved for the final test and not used during the training of ACGAN. Then, part of the fault data in Data 2 (shown in the red dashed box) is removed deliberately to mimic a situation of imbalanced data encountered in practical scenarios. Afterwards, the remaining small amount of fault data is used to train the ACGAN, specifically for its generator G. The data highlighted by black bars in the flowchart represent fault data generated by the generator, and its quantity is equivalent to the amount of data removed from the red dashed box. Such synthetic data are subsequently fused with the imbalanced data to construct new data for the training of the classification model, i.e., 1DCNN. Finally, the performance of the trained classification model is tested on the separate Data 1.

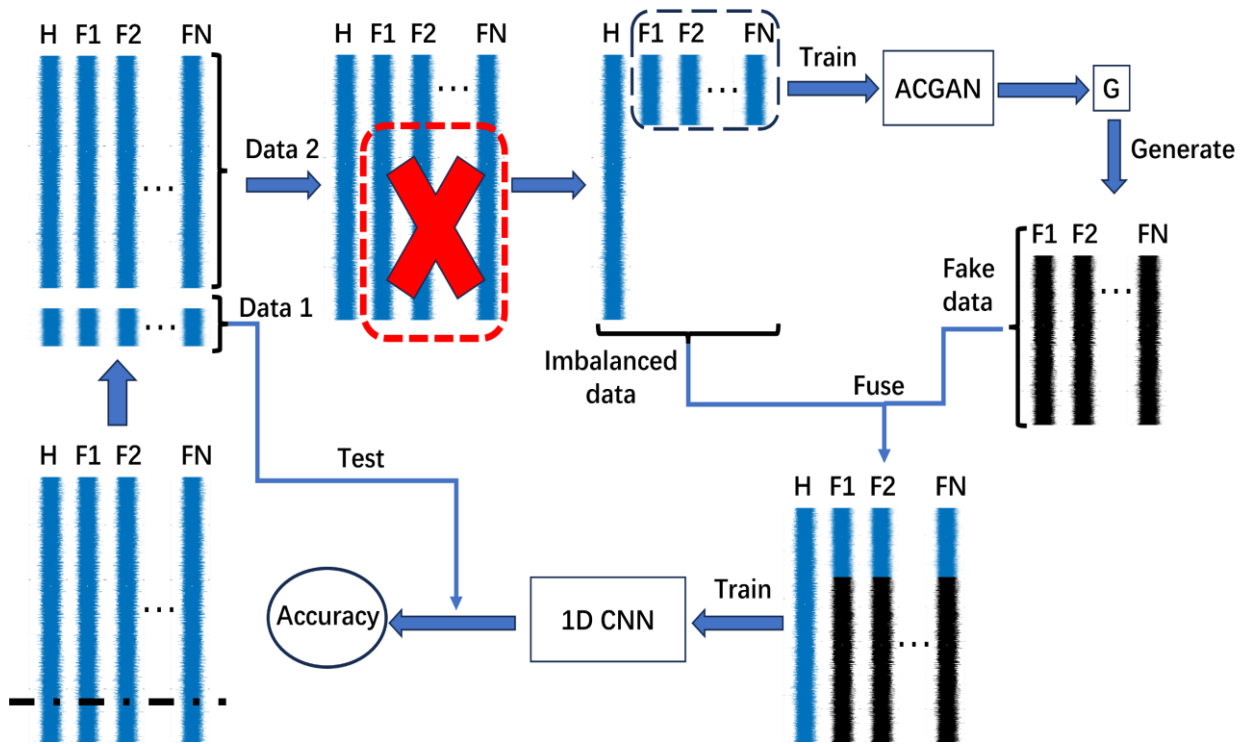


Figure 38. The flowchart of proposed fault diagnosis method.

5.3 Experiments and Results

5.3.1 Experiments on the Time-Varying Rotational Speed Bearing Data

5.3.1.1 A Brief Introduction to the Data and Modelling Tasks

The Time-Varying Rotational Speed (TVRS) bearing data can be found in [178], which were collected from a SpectraQuest machinery fault simulator. The dataset consists of three different states of data: healthy state, faulty with an inner race defect, and faulty with an outer race defect. For each state, there are four different speed conditions, namely, increasing speed, decreasing speed, increasing then decreasing speed and decreasing then increasing speed. For each experimental setting, three trials were carried out to ensure the authenticity of the data, generating 36 datasets in total. Some detailed information of the data is shown in Table 26 and the presentation of raw data is shown in Figure 39. In Section 5.3.1, only H1, F1, and F2 data are used to test the ability of the proposed method for the identification and classification of two types of faults, which occur within 1s, 2s, 3s, and 4s. The reason is to simulate the scenario that new WTs have a relatively shorter operating time and have not experienced many types of malfunctions. All these data are sampled with 200k Hz, as defined by the specifications of the publicly available TVRS bearing dataset. Although in many researches, a sampling rate of 10–15 kHz is considered sufficient to capture bearing fault characteristics, the higher sampling rate allows for the preservation of more detailed high-frequency features which can be beneficial for DL-based diagnosis. As a control experimental group, a dataset sampled with 12kHz is employed in Section 5.3.2.

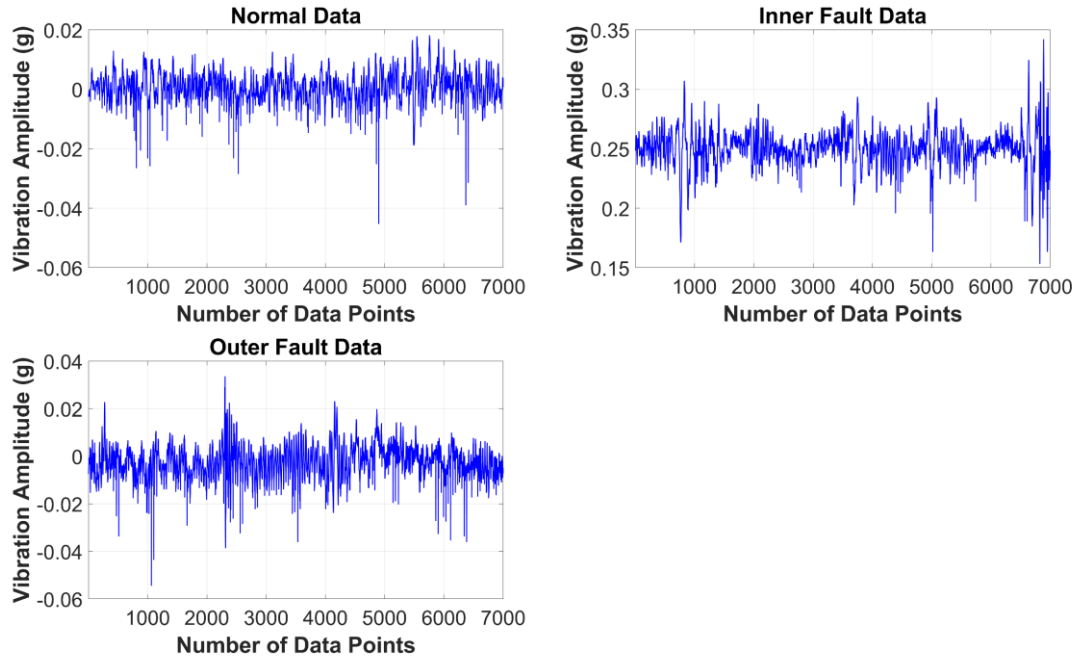


Figure 39. Presentation of raw data in TVRS dataset.

Table 26. Details of TVRS bearing data.

Speed \ Fault	Healthy state	Inner fault	Outer fault
Increasing speed	H1	F1	F2
Decreasing speed	H2	F3	F4
Increasing then decreasing speed	H3	F5	F6
Decreasing then increasing speed	H4	F7	F8

As shown in Figure 38, the three types of data, one health and two types of fault data, are divided into two parts: Data1 (10%) and Data2 (90%), as described in the last paragraph of Section 5.2.2. Data 1 is reserved for final test; part of the Data 2 is removed deliberately to mimic a situation of imbalanced data encountered in real practice. The remaining fault data in Data 2 is employed to train a GAN, ACGAN and Multi-Scale ACGAN, respectively. Note that other schemes have also been used to create an imbalanced data scenario, e.g. in [179, 180], it created a situation by changing the ratio of fault data to health data. A primary limitation of this scheme is that in real practice machines are not allowed to run in fault for a long time, and usually the quantity of fault data is very limited. In [179, 180], it paid much attention to the ratio of health data to fault data but ignored the quantity

of fault data. In reality, the lack of fault data is usually a more changing issue than the ratio of fault to health data. Therefore, in this work experiments are carried out under different fault durations, 1s, 2s, 3s, and 4s. Note that the sampling frequency of TVRS is 200 kHz, making the number of fault data points be 200k, 400k, 600k, and 800k, respectively. For comparative purposes, GAN and ACGAN are also employed to generate fault data. The training process of GAN, ACGAN, and Multi-Scale ACGAN are shown in Figure 40, Figure 41, and Figure 42, respectively.

5.3.1.2 Training Process of GAN on TVRS Bearing Data

In Figure 40, the top panel demonstrates the loss curves of the generator and discriminator over epochs. The generator loss (blue line) begins at a high value, which is typical at the start of training when the generator's outputs are far from the real data distribution. The discriminator loss (orange line) starts at a lower value but exhibits significant fluctuations in the early epochs, indicating that the discriminator learns to distinguish between real and generated data. Both the generator and discriminator losses stabilize after approximately 700 epochs, although both still consistently oscillate with a small amplitude. This behavior suggests that the generator and discriminator reach a state of equilibrium where neither is improving significantly over the other. However, small fluctuations persist throughout the training process, reflecting the ongoing adversarial nature of GAN training.

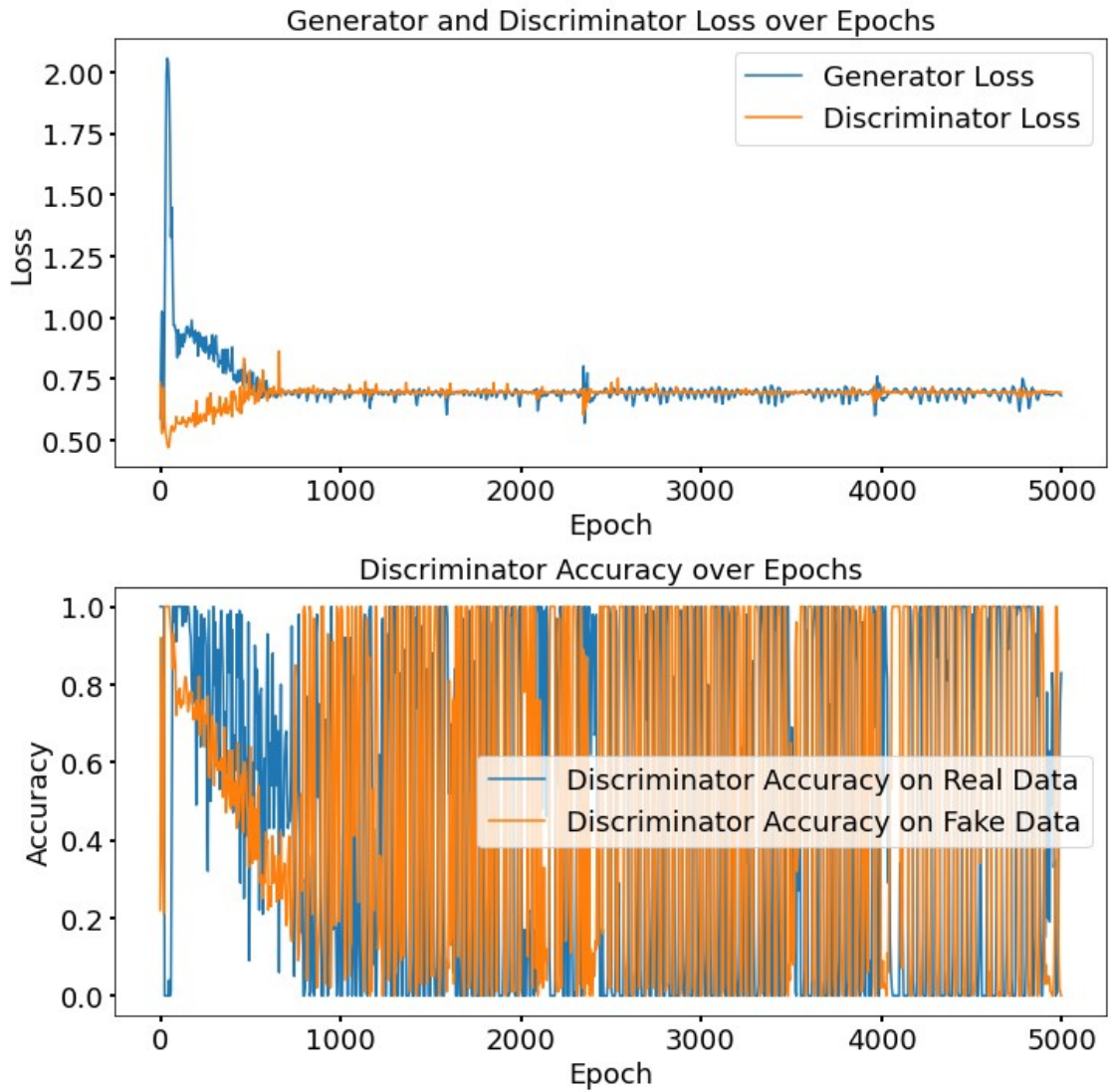


Figure 40. The training process of GAN on TVRS bearing data.

The bottom panel in Figure 40 illustrates the discriminator's accuracy on both real (blue line) and generated (fake) data (orange line) throughout the training process. Initially, the discriminator exhibits high accuracy on both real and fake data, which suggests that it effectively distinguishes between the two classes. As the training progresses, there is a marked drop in accuracy, especially on the fake data, implying that the generator is beginning to produce more realistic samples, thus making the discriminator's task more challenging. After approximately 900 epochs, the discriminator's accuracy on both real and fake data becomes highly unstable, reflecting the adversarial dynamics where both the generator and discriminator are continuously improving. The near-random oscillations between high and low accuracy values in later epochs suggest that the GAN is experiencing a form of mode collapse or cycling behavior, where the generator occasionally produces a limited diversity of samples that the discriminator either identifies or fails.

5.3.1.3 Training Process of ACGAN on TVRS Bearing Data

The top panel in Figure 41 illustrates the progression of the generator and discriminator losses over the course of 5,000 training epochs. The generator's loss (blue line) exhibits a general trend of stabilization after initial fluctuations. In the early stage of training, there are obvious increases and decreases, suggesting moments where the generator tried to improve its generation ability. The loss becomes more consistent as the training progresses, indicating that the generator learns to produce more realistic data. The loss of the discriminator (orange line) is initially at a similar level to that of the generator. As training progresses, it shows a downward trend and eventually stabilizes. The loss eventually stabilizes at a lower level, although it exhibits occasional spikes, possibly due to changes in the generator's strategy. The decreasing and stabilizing trend in both the generator and discriminator losses suggests that the adversarial training is balancing out, with neither model consistently outperforming the other. This indicates that the model is reaching an equilibrium where the generator is producing more realistic samples, and the discriminator is appropriately challenged by these samples.

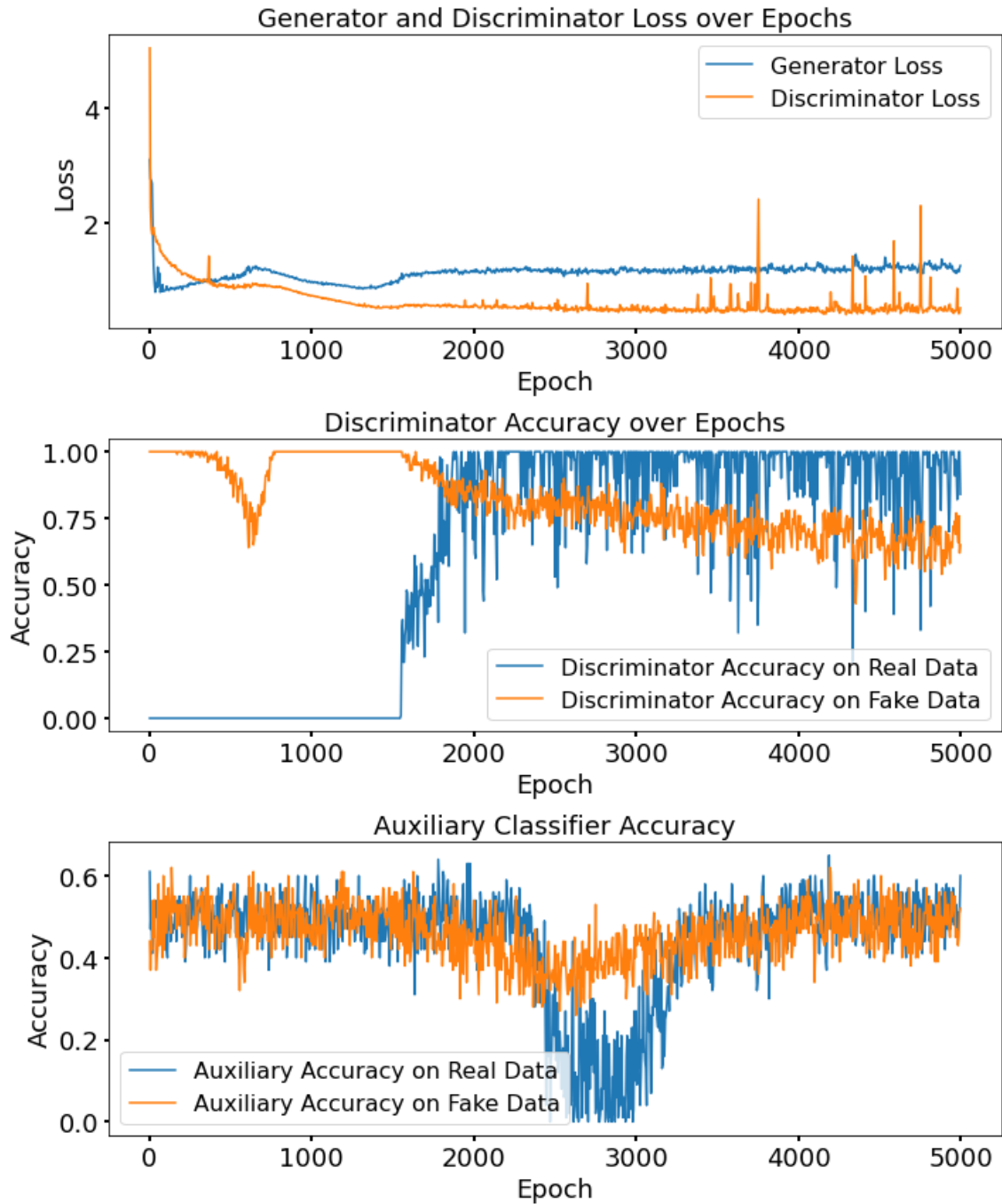


Figure 41. The training process of ACGAN on TVRS bearing data.

The middle panel in Figure 41 shows the accuracy of the discriminator on both real and fake data. At the beginning of the training, the discriminator does not work for the real data (blue line) at all, with accuracy being close zero, indicating that the discriminator initially struggles to correctly identify real data, possibly due to being overly biased towards predicting data as fake. Conversely, the discriminator works perfect for the fake data (orange line), with accuracy being close to 100%, but the accuracy decreases rapidly after around 300 epochs, reflecting the generator's initial success in producing more realistic fake samples that are challenging for the discriminator. Subsequently, the

accuracy of fake data increases rapidly and stabilizes for a period before experiencing another decline, indicating that the generator tries to improve and produce more convincing fake samples but failed. After around 1,900 epochs, the discriminator's accuracy on real data quickly rises to 100%, suggesting that the discriminator adapts and becomes proficient at recognizing real data after an initial period of adjustment. In the later stages of training, both accuracy curves exhibit substantial oscillations, reflecting the ongoing adversarial nature of the GAN, where the generator and discriminator continuously refine their strategies against each other.

The final panel in Figure 41 illustrates the accuracy of the auxiliary classifier on both real data (blue line) and fake data (orange line) across the 5000 epochs. The accuracy of the auxiliary classifier on both real and fake data hovers around 50% throughout the training process, with significant fluctuations. This relatively low and fluctuating accuracy suggests that the auxiliary classifier is struggling to consistently differentiate between the classes. The persistence of these fluctuations indicates that the auxiliary classifier has not converged effectively, which may necessitate further investigation or refinement of the model architecture or training strategy.

5.3.1.4 Training Process of Multi-Scale ACGAN on TVRS Bearing Data

Figure 42 shows the training process of Multi-Scale ACGAN on the TVRS bearing data. The first panel in Figure 42 shows the loss values for both generator and discriminator as the training progresses. The generator loss (blue curve) initially decreases sharply during the first 200 epochs, indicating that the generator is quickly improving its ability to generate realistic samples. The discriminator loss (orange curve) remains relatively stable after an initial slight decrease, suggesting that the discriminator is effectively learning to distinguish between real and generated data. As the training progresses, the discriminator loss increases slightly and stabilizes around the same level as the generator's loss. This indicates a balance between the generator and discriminator, which is characteristic of a GAN reaching equilibrium.

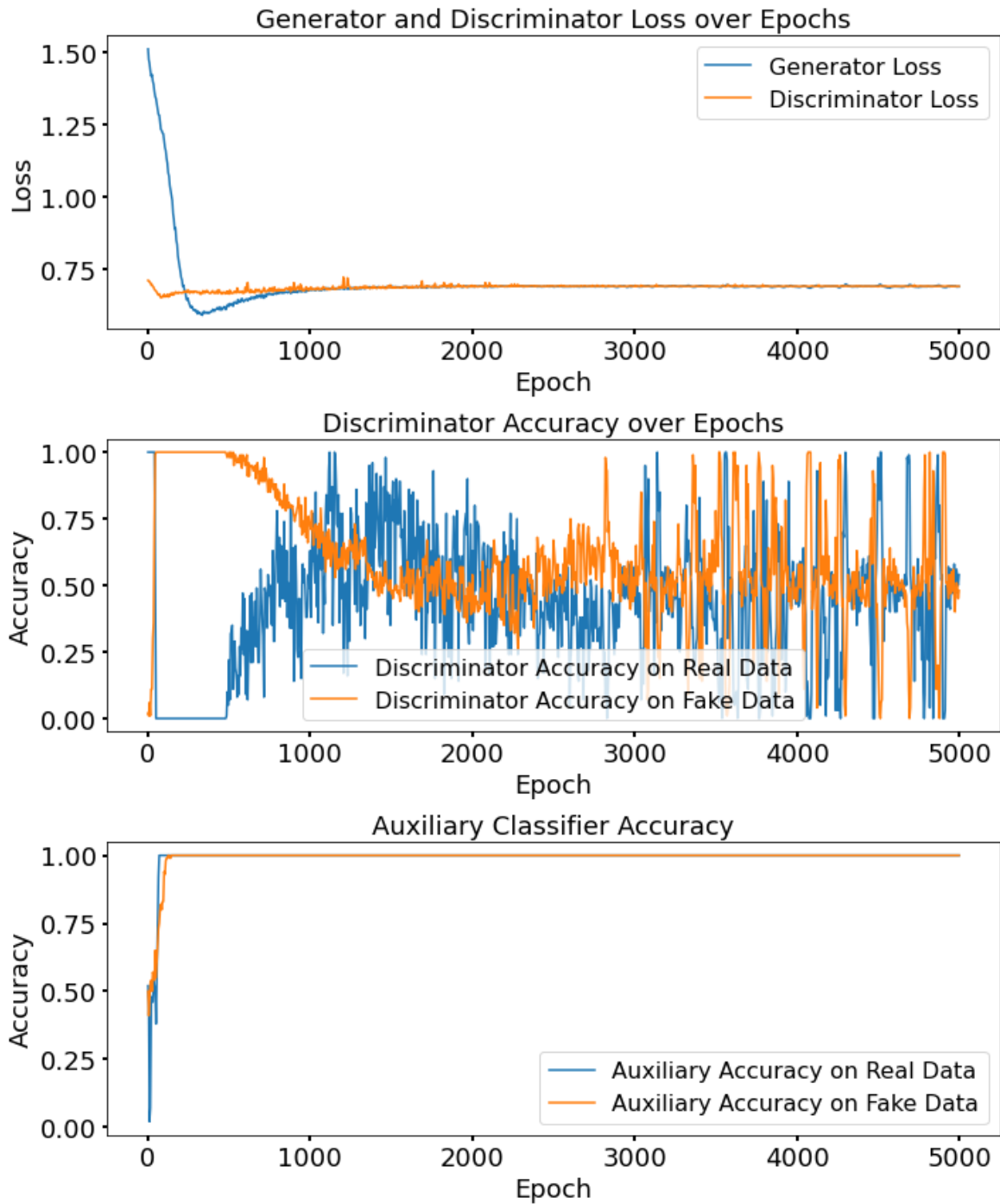


Figure 42. The training process of Multi-Scale ACGAN on TVRS bearing data.

The second panel in Figure 42 shows the accuracy of the discriminator in distinguishing real data (blue curve) and fake data (orange curve). At the early stage of training, the generator is still in its early learning phase, producing low-quality synthetic data that the discriminator can easily distinguish from real data. As a result, the discriminator works well for identifying the fake data and thus achieves an accuracy close to 100%. The discriminator has a poor ability to classify real data at the early stage, probably because the randomly initialized weights have not yet enabled it to effectively distinguish

real data. As a result, in the first few iterations of training, the discriminator will misclassify a large amount of real data as fake data, causing its accuracy on real data to be close to 0. However, as the training progresses, the accuracy of the discriminator in identifying real data gradually increases, while the accuracy of identifying fake data gradually decreases, and finally, the two reach a balance, almost stabilizing at 50%. This is a good indication that after a long period of training, the generator and the discriminator conflict with each other, making it difficult for the discriminator to distinguish the authenticity of the data.

The third panel in Figure 42 displays the accuracy of the auxiliary classifier, which predicts the class labels of the data, for both real and fake data. The auxiliary classifier's accuracy on both real and fake data (blue and orange lines, respectively) increases quickly and stabilizes near 100%, suggesting that the classifier is highly effective at predicting the correct class labels. This rapid convergence to high accuracy indicates that the auxiliary classifier successfully learns to classify the data, which is critical for the proposed Multi-Scale ACGAN's performance.

5.3.1.5 Performance of Different Fault Diagnosis Methods on TVRS Bearing Data

As shown in Figure 38, the generated faulty data and imbalanced data are combined to create artificial balanced data. The classification model is then trained with this data, and the classification test is performed on the originally reserved Data 1 (see the last paragraph of Section 5.2.2). To detect and inspect the impact of different quantities of fault data on the final accuracy of fault classification, experiments were conducted under four different fault duration scenarios labelled as 1s, 2s, 3s, and 4s. Note that the sampling rate of TVRS bearing data is 200kHz, so the lengths of fault data are 200k, 400k, 600k, and 800k, respectively. The experimental results are shown in Figure 43.

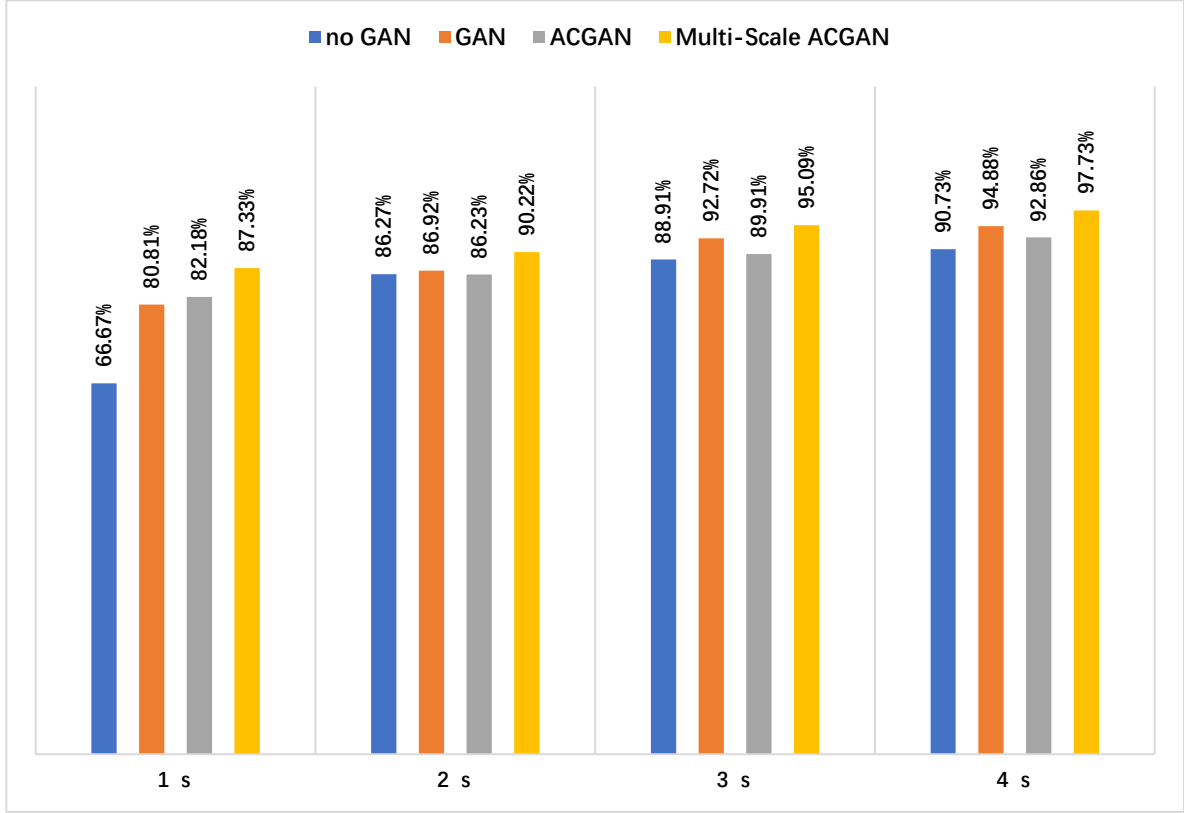


Figure 43. The fault diagnosis accuracy on TVRS bearing data with different methods for the four fault durations: 1 s, 2 s, 3 s and 4 s.

The results demonstrate the advantages of using GAN-based methods to deal with imbalanced data scenarios. Specifically, the no GAN condition, which relies solely on a limited and imbalanced dataset, yields the lowest accuracy across all scenarios, with the highest being 90.73% in the 4S scenario. In contrast, the introduction of GAN improves accuracy, with the Multi-Scale ACGAN method consistently achieving the highest accuracy across all scenarios. For instance, in the 1S scenario, Multi-Scale ACGAN reaches 87.33% accuracy, significantly outperforming the no GAN method (66.67%). As the quantity of fault data increases, Multi-Scale ACGAN maintains its superiority, achieving accuracies of 90.22%, 95.09%, and 97.73%, respectively. These results suggest that the use of GAN-based methods, particularly Multi-Scale ACGAN, substantially enhances the fault diagnosis performance by effectively mitigating data imbalance and enriching the dataset with synthetic but realistic fault data.

Except for the structure of the generator and discriminator, activation functions can also have a great influence on the performance of GAN-based fault diagnosis methods. The Linear activation function can produce continuous data and will not limit the amplitude of produced data, so the Linear

activation function for the generator is used as one control group. The activation function condition is noted as C1. In [112], the activation function of the generator is Tanh, and there is no activation function for the discriminator, but in our research, the input data are vibration signals which are different from the 2D images. Thus, for comparative purposes, the proposed Multi-Scale ACGAN is changed: Tanh is chosen as the activation function of the generator, and no activation function is used for the discriminator. This condition is conducted as another control group and noted as C2. The proposed activation function condition, shown in Table 24 and Table 25, is noted as C3. Details of three activation function scenarios are recorded in Table 27. The results of the proposed Multi-Scale ACGAN with different activation functions are shown in Figure 44.

Table 27. The details of three activation functions.

Activation function scenarios	Discriminator	Generator
C1	Sigmoid and SoftMax	Linear
C2	None	Tanh
C3	Sigmoid and SoftMax	Tanh

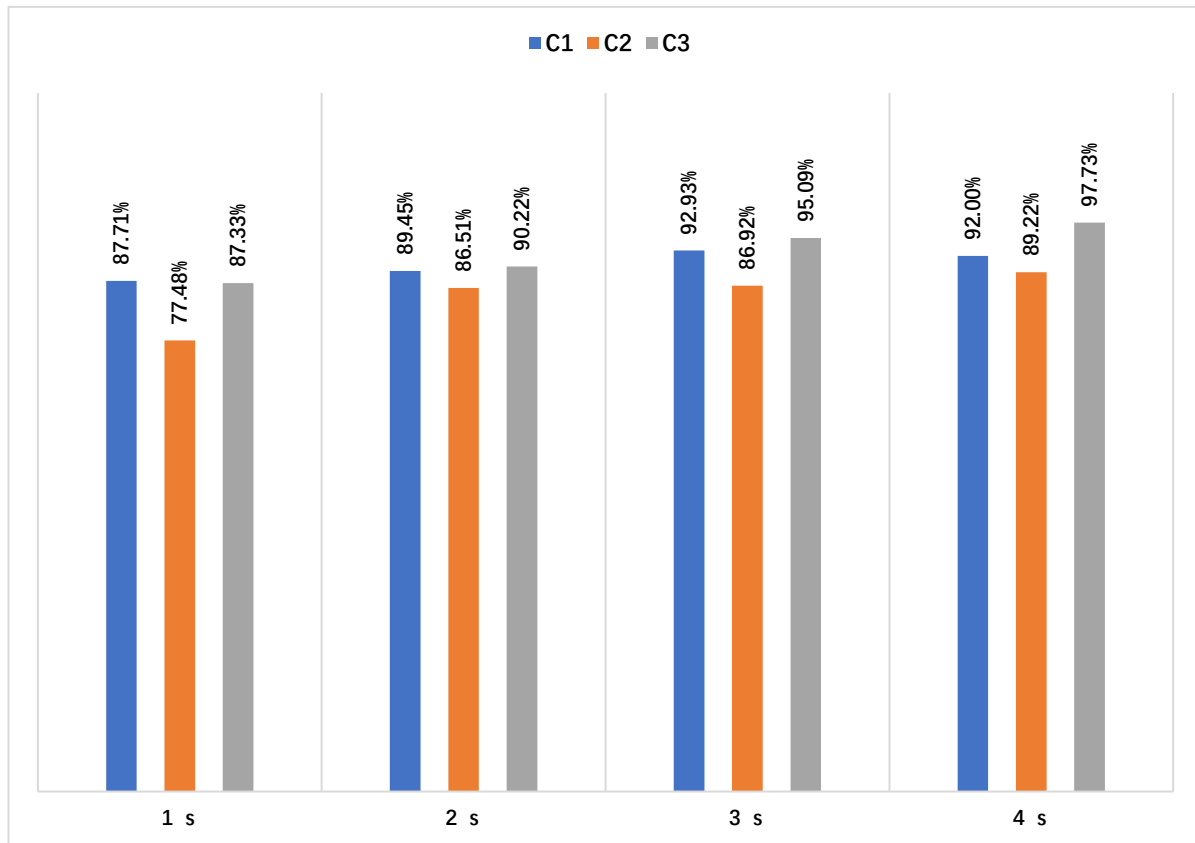


Figure 44. The fault diagnosis accuracy of Multi-Scale ACGAN on TVRS bearing data with different activation functions for the four fault durations: 1 s, 2 s, 3 s and 4 s.

Across all four scenarios, condition C3, the proposed Multi-Scale ACGAN, consistently achieves the highest accuracy rates, indicating it is the most effective and reliable method. Condition C1 also shows strong performance but is generally outperformed by C3. Condition C2 consistently shows lower accuracy compared to the other two, suggesting it is less effective for fault diagnosis under these conditions. The overall trend suggests that condition C3 is the most robust and accurate method across various scenarios. When the fault data can be collected within 4s, the fault diagnosis accuracy of the proposed method can reach up to 97.73%.

5.3.1.6 A Brief Summary for Experiments on TVRS Bearing Data

In summary, the results conducted on the TVRS bearing data demonstrate that the designed Multi-Scale ACGAN can significantly improve the performance of fault diagnosis under imbalanced data conditions. From Figure 40 and Figure 41, when the traditional GAN and ACGAN were applied to the imbalanced data, they do not work well to reach equilibrium but the designed Multi-Scale ACGAN can reach the equilibrium state as the training progress goes, as shown in Figure 42. In addition, with the help of GAN, the fault diagnosis accuracy can be significantly increased compared to without using GAN.

5.3.2 Experiments on the Case Western Reserve University Bearing Data

5.3.2.1 A Brief Introduction to the Data and Modelling Tasks

In Section 5.3.1, there are only two types of fault data, but now, the experiments are carried out on the CWRU bearing data [169], and the types of fault data are increased to nine. The sampling frequency of CWRU is 12 kHz and only the drive end accelerometer data are used. There are three different fault positions: ball fault, inner fault, and outer fault. In addition, there are four workloads, 0 HP, 1HP, 2HP, and 3HP. However, in Section 5.3.2, only the data collected under 0HP is used. Apart from the fault positions and workloads, there are three different fault diameters, 0.18mm, 0.36mm, and 0.53mm. In summary, there are one healthy state and nine fault states in this dataset. Some details of the CWRU bearing data are listed in Table 28.

Table 28. Details of CWRU bearing data.

Health condition type	Label classes
Normal	H
BF-0.18	F1
BF-0.36	F2
BF-0.54	F3
IF-0.18	F4
IF-0.36	F5
IF-0.54	F6
OF-0.18	F7
OF-0.36	F8
OF-0.54	F9

Similar to the experimental settings in Section 5.3.1, faults occurred within 1s, 2s, 3s, and 4s. Note that the sampling frequency of CWRU is 12 kHz, making the number of fault data points be 12k, 24k, 36k, and 48k, respectively. For comparative purposes, GAN and ACGAN are also employed to generate fault data.

5.3.2.2 Training Process of GAN on CWRU Bearing Data

The training process of GAN is shown in Figure 45. In the top panel of the figure, the blue line represents the generator's loss over the training epochs. Initially, the generator loss is quite high, reaching up to approximately 2.2. This is typical in early training phases, where the generator still learns to produce data that can deceive the discriminator. As training progresses, the generator loss fluctuates violently but generally trends downward, indicating that the generator improves its ability to produce realistic data. Conversely, the discriminator loss (orange line) starts relatively low, indicating initial effectiveness at distinguishing real data from fake. However, as training progresses, the loss experiences fluctuations, slightly trending upwards. This trend suggests an increasing challenge for the discriminator to accurately classify the real and generated data as the generator performance improves. The stabilization but with slight fluctuations observed indicates that the

discriminator adjusts to new patterns learned by the generator, thereby reflecting the ongoing adversarial competition between the two components.

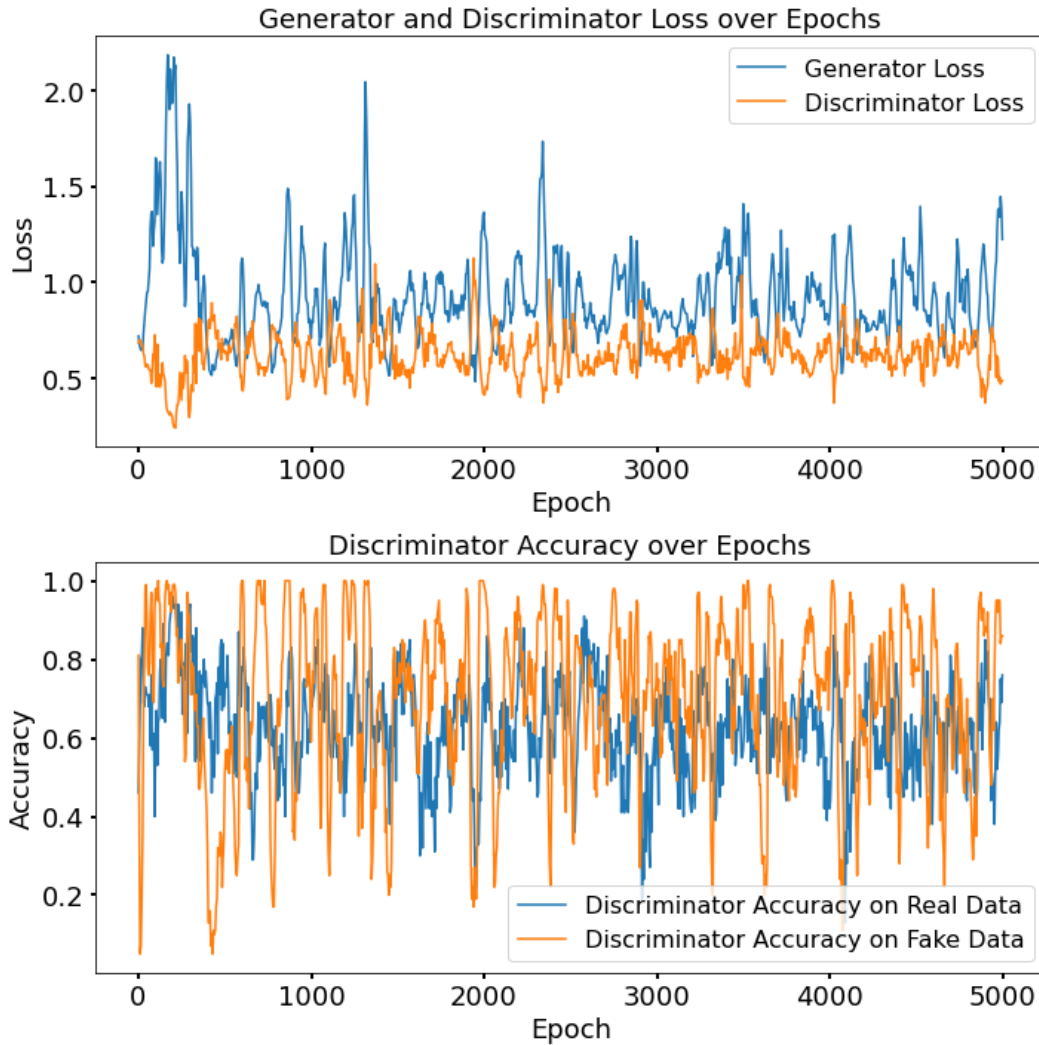


Figure 45. The training process of GAN on CWRU bearing data.

The bottom graph in Figure 45 reflects the discriminator's ability to distinguish between real data (blue line) and fake data (orange line) throughout the training process. At the beginning, the discriminator has a high recognition rate for both real and fake data, which reflects that the generator does not perform well to generate realistic data to confuse the discriminator. However, as the training progresses, the discriminator's accuracy in distinguishing between real and fake data decreases and slowly fluctuates within a larger range. This means that the generator becomes more able to generate realistic data, making it difficult for the discriminator to distinguish between real and fake data. However, the accuracy that the discriminator correctly distinguishes the data, especially fake data,

has been oscillating within a large range, showing that the generator and the discriminator engage in continuous adversarial training, making it difficult to reach a relatively balanced state.

5.3.2.3 Training Process of ACGAN on CWRU Bearing Data

Figure 46 shows the training process of ACGAN on the CWRU bearing data. The top panel in the figure shows the loss curves of the generator and discriminator of ACGAN over 5,000 epochs. The generator loss (blue line) decreases rapidly during the initial stages of training, indicating that the generator quickly improves in producing data that can fool the discriminator. Over time, the loss continues to decrease gradually, stabilizing with minor fluctuations towards the latter part of the training. This suggests that the generator learns to produce increasingly realistic samples. The discriminator loss shows a more stable trend throughout the training process. Initially, it decreases but eventually stabilizes around a low value, indicating that the discriminator effectively learns to differentiate between real and generated data. In the final stage, both losses appear to converge at a low level. This suggests that the model training is approaching a point where the generator and discriminator are in a balanced state, which is a characteristic of adversarial networks, where neither model significantly outperforms the other.

The middle panel shows the fault diagnosis accuracy of the discriminator on both real and fake data. The orange line represents the accuracy of the discriminator in identifying fake data, which is initially very high. This indicates that at the beginning of the training, the generator can hardly produce data that closely resembles real data, making it easy for the discriminator to identify fake data. However, as training progresses, accuracy fluctuates and shows an overall downward trend. This suggests that the generator's ability to produce realistic data improves, making it harder for the discriminator to distinguish between real and fake data. The blue line represents the accuracy of the discriminator in identifying real data. Initially, the accuracy is low, but it gradually increases as training progresses. This indicates that the discriminator's ability to differentiate is enhanced over time. In the final stages, the accuracy of the discriminator in identifying both fake and real data stabilizes around 65% with fluctuations. This indicates that, after training, the discriminator finds it challenging to distinguish between real and fake data. This outcome reflects significant improvements in the generator's capability, suggesting that the ACGAN has reached a stable state.

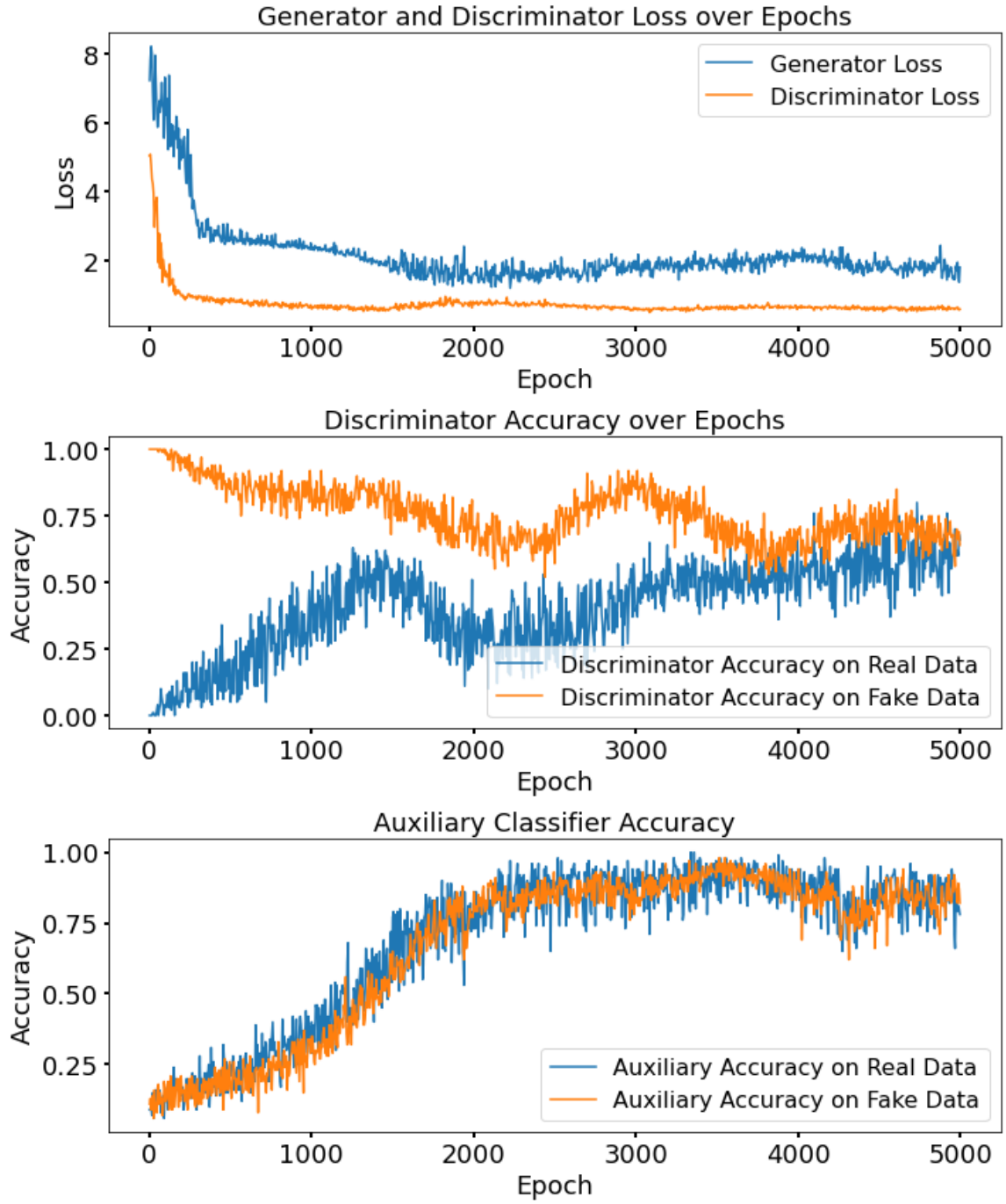


Figure 46. The training process of ACGAN on CWRU bearing data.

The bottom panel shows the accuracy of the auxiliary classifier when classifying the class labels of real and fake data. The accuracy of the auxiliary classifier on both real data and fake improves over time and eventually stabilized at high values, suggesting that the model is effective in generating samples that not only resemble the real data but also contain meaningful class information.

5.3.2.4 Training Process of Multi-Scale ACGAN on CWRU Bearing Data

Figure 47 shows the training process of Multi-Scale ACGAN for the CWRU bearing data. The first panel in the figure shows the loss curves of the generator and discriminator of the proposed Multi-Scale ACGAN. At the beginning of training, the generator's loss (blue curve) starts relatively high, indicating that the generator initially struggles to produce realistic data. The discriminator's loss (orange curve), on the other hand, is relatively low, suggesting that it easily distinguishes between real and generated data. As training progresses, the generator's loss sharply decreases, stabilizing around a lower level after approximately 1000 epochs. This indicates that the generator improves its ability to produce data that can fool the discriminator. Meanwhile, the discriminator's loss remains low and relatively stable, which is typical as it continues to successfully differentiate between real and fake data, albeit with more difficulty as the generator improves. Toward the end of training (after 4000 epochs), both losses stabilize, indicating that the model has reached a relatively balanced state. The losses remaining low suggest that the adversarial process between the generator and discriminator has reached a point where neither model is significantly outperforming the other, suggesting convergence.

The second panel in Figure 47 presents the accuracy of the discriminator when classifying real and fake data over the course of training. In the early epochs, the discriminator accuracy shows significant fluctuations for both real (blue curve) and fake data (orange curve). Such an instability is common during the early phases of ACGAN training due to the dynamic adversarial process. As training progresses, these accuracies fluctuate around 50%, suggesting that the discriminator finds it increasingly difficult to distinguish between real and fake data, which is an expected outcome as the generator improves. The persistent fluctuations throughout training indicate a continual back-and-forth between the generator and discriminator, where each improves in response to the other. The lack of a clear trend towards 0% or 100% accuracy for either real or fake data suggests that the model maintains a healthy adversarial balance without either network overpowering the other.

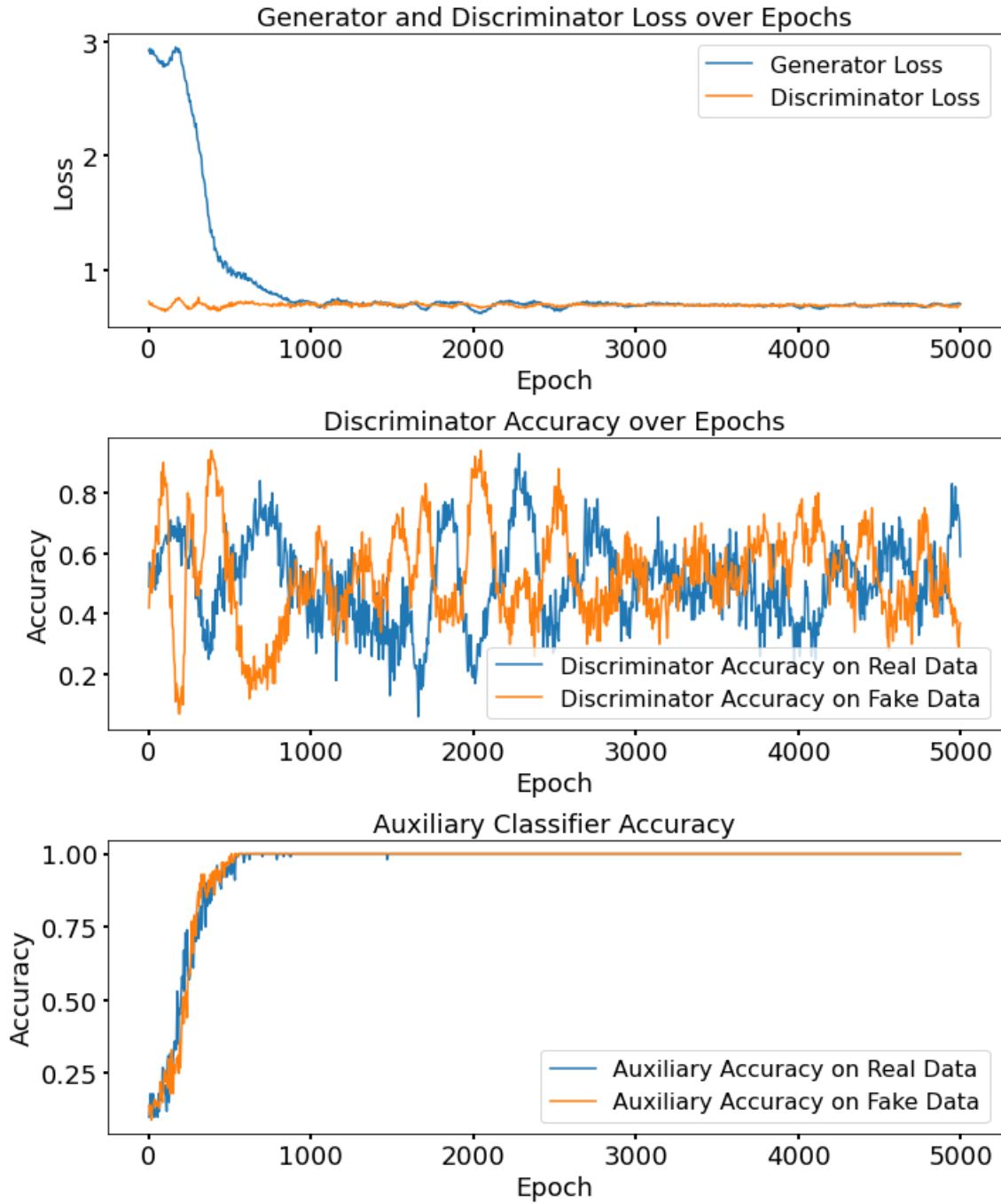


Figure 47. The training process of Multi-Scale ACGAN on CWRU bearing data.

The third panel in Figure 47 illustrates the accuracy of the auxiliary classifier, which is tasked with classifying the correct labels of both real and generated data. The accuracy of the auxiliary classifier on both real (blue curve) and fake data (orange curve) shows rapid improvement, quickly converging to near-perfect accuracy (close to 100%) within the first 1000 epochs. This indicates that the auxiliary classifier efficiently learns to identify the correct labels for both real and generated data. After reaching high accuracy, the classifier maintains this performance throughout the remaining

epochs. The convergence to high accuracy and its stability suggests that the auxiliary classifier is well-calibrated and consistently reliable in its predictions.

The overall training process demonstrates typical and desirable characteristics of a well-trained ACGAN. The generator and discriminator exhibit a healthy adversarial relationship, as evidenced by the convergence of their losses and the fluctuating discriminator accuracies. The auxiliary classifier performs exceptionally well, achieving high accuracy early in the training and maintaining it throughout. This outcome indicates a successful multi-scale learning within the proposed Multi-Scale ACGAN framework.

5.3.2.5 Performance of Different Fault Diagnosis Methods on CWRU Bearing Data

Following the procedure shown in Figure 38, the generator is employed to generate faulty data once the Multi-Scale ACGAN is well trained. Then, the generated fault data and imbalanced data are combined to create artificial balanced data. The classification model is an 1DCNN, which is then trained with the data, and the classification test is performed on the originally reserved Data 1 (see the last paragraph of Section 5.2.2). Similarly, to detect and inspect the impact of different quantities of fault data on the final accuracy of fault classification, experiments on CWRU bearing data were also conducted under four different fault duration scenarios labelled as 1s, 2s, 3s, and 4s. However, the sampling rate of CWRU bearing data is 12 kHz, so the lengths of fault data are 12k, 24k, 36k, and 48k, respectively. The experimental results are shown in Figure 48.

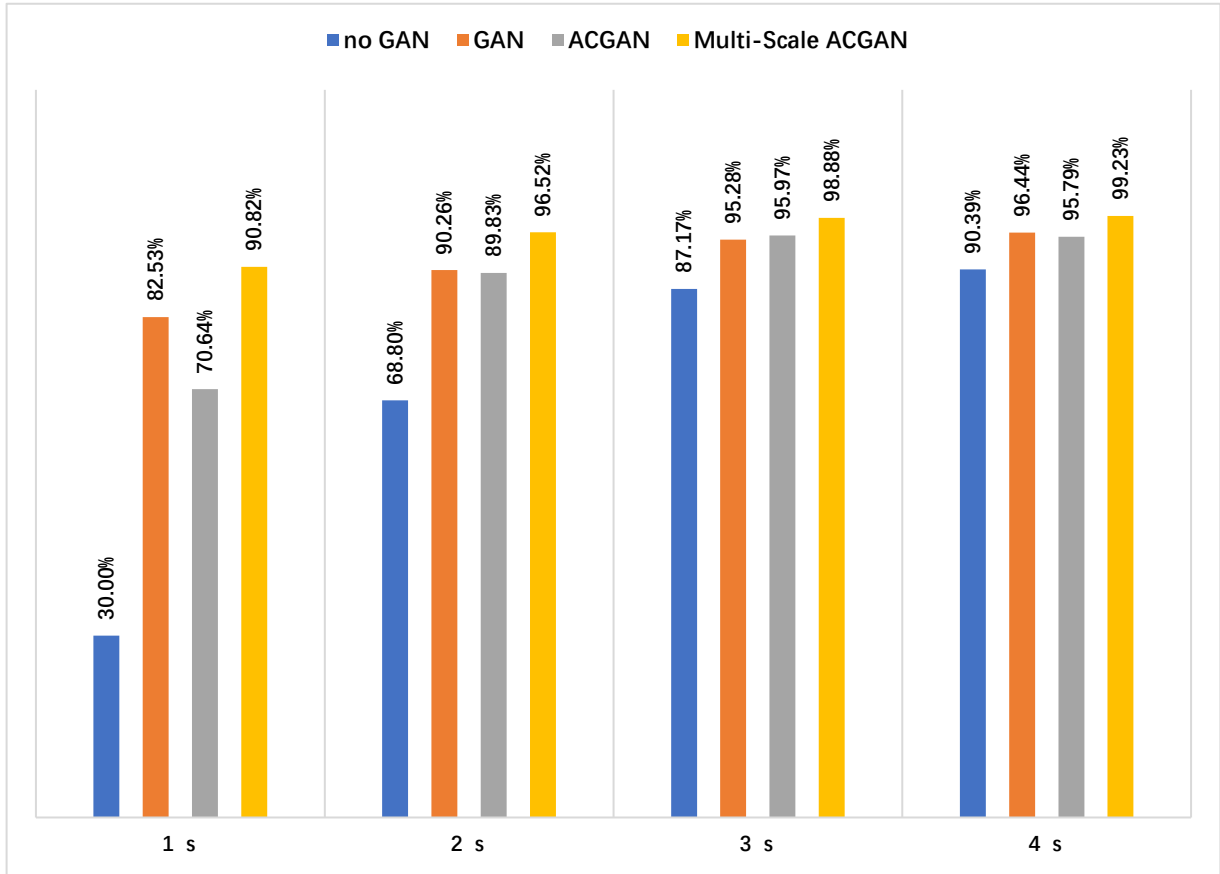


Figure 48. The fault diagnosis accuracy on CWRU bearing data with different methods for the four fault durations: 1 s, 2 s, 3 s and 4 s.

The results clearly show that the direct use of the imbalanced dataset (without using GAN) yields the lowest accuracy across all time scenarios, particularly for scenario with the fault duration of 1 s, where it only achieves an accuracy of 30%. When the GAN-based methods are applied, a significant improvement in accuracy is achieved. The Multi-Scale ACGAN method consistently outperforms the other methods, achieving the highest accuracy for each fault scenario, with an impressive accuracy of 90.82% for the 1 s case, and 99.23% for the case of 4 s. This indicates that as the quantity of the collected data increases, the effectiveness of the GAN-based methods, particularly the Multi-Scale ACGAN, becomes even more pronounced. The Multi-Scale ACGAN's superior performance highlights its robustness and reliability in addressing the challenges posed by imbalanced datasets in bearing fault classification, making it the most effective method among those evaluated.

To find out the influence of activation functions on the performance of fault diagnosis methods, experiments with three different activation functions, described in Table 27, are carried out. The results of the proposed Multi-Scale ACGAN with the three different activation functions are shown in Figure 49.

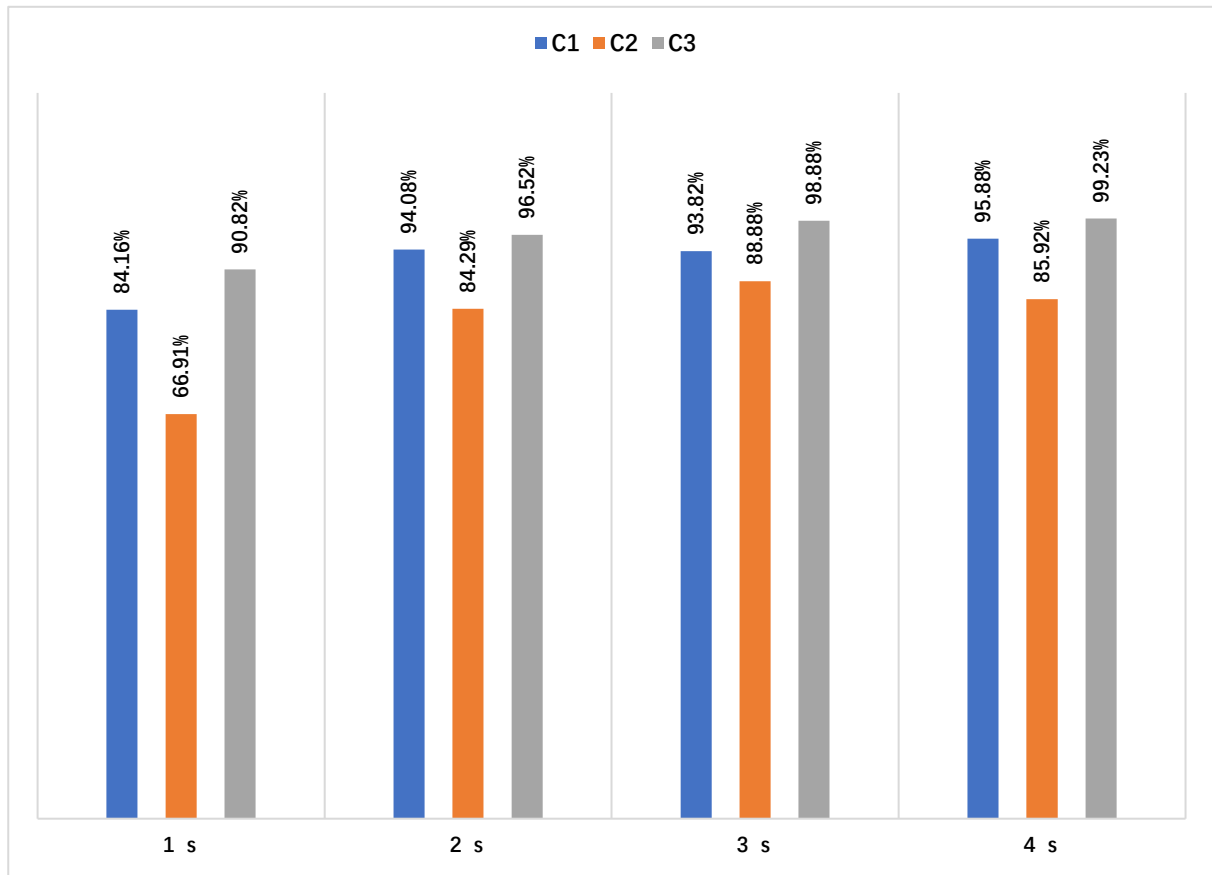


Figure 49. The fault diagnosis accuracy of Multi-Scale ACGAN on CWRU bearing data with different activation functions for the four fault durations: 1 s, 2 s, 3 s and 4 s.

The results indicate that the choice of activation function significantly impacts the model's overall performance. For all the four fault cases (durations), the activation function C3 consistently achieves the highest accuracy, demonstrating its superiority over C1 and C2. Specifically, C3 achieves an accuracy of 90.82% in the 1 s case, notably higher than 84.16% achieved by using C1 and 66.91% achieved by using C2. Similar observations are noted for the case of 2 s fault type, where C3 reaches an accuracy of 96.52%, surpassing C1 and C2 with 94.08% and 84.29%, respectively. For the other two cases of 3 s and 4 s, C3 maintains its lead with accuracies of 98.88% and 99.23%, compared to 93.82% and 95.88% achieved by C1, and 88.88% and 85.92% achieved by C2. These results underscore the robustness of the activation function C3, which consistently delivers higher accuracy, particularly in more fault data scenarios, making it the most reliable choice among the three tested functions.

5.3.2.6 A Brief Summary for Experiments on CWRU Bearing Data

Experimental results on the CWRU bearing data demonstrate that the fault diagnosis accuracy can be significantly increased by introducing GAN. The designed Multi-Scale ACGAN outperforms the traditional GAN and ACGAN for solving imbalanced data classification tasks. From Figure 45, Figure 46 and Figure 47, it is obvious that the loss curve and the accuracy curve of the auxiliary classifier of the Multi-Scale ACGAN are the smoothest among the three fault diagnosis methods. Additionally, compared to the traditional ACGAN, the auxiliary classification accuracy curve of the Multi-Scale ACGAN is also more stable.

5.4 Discussion about the Experimental Results of the Proposed Fault Diagnosis Method

The experimental results presented in Section 5.3 validate the superior performance of the proposed Multi-Scale ACGAN method in handling data imbalance fault diagnosis tasks. Compared to the original GAN and conventional ACGAN, the Multi-Scale ACGAN consistently achieved higher accuracy across both the TVRS and CWRU datasets. This improvement is particularly evident in cases with short fault durations (1 s and 2 s), where the available real fault samples are extremely limited.

The effectiveness of the proposed method can be attributed to two key enhancements. Firstly, the multi-scale parallel feature extraction structure allows the generator and discriminator to better capture features across different scales, thereby improving the quality of synthetic samples. Secondly, the integration of class labels in both the generator and discriminator ensures that generated samples are not only realistic but also class-consistent, improving the classification performance. Moreover, the comparative experiments using different activation functions further confirm the robustness of the proposed model.

In summary, the results demonstrate that the proposed method not only addresses the data imbalance issue effectively but also generalizes well across datasets and fault durations. This makes it a promising approach for practical industrial applications where data imbalance is common and early diagnosis is critical.

5.5 Conclusion about the Proposed Multi-Scale ACGAN-Based Fault Diagnosis Method

In this chapter, a novel Multi-Scale ACGAN-based bearing fault diagnosis method is proposed to solve classification tasks with imbalanced data. The existing generator and discriminator structures were adapted, and new generator and discriminator models were developed. The impact of the activation functions on the overall performance of the classifiers was analyzed. It is expected that the analysis results are useful for properly choosing an activation function.

The performance of the proposed method is verified on two public datasets: TVRS and CWRU, both have several different types of faults. The experimental results demonstrate that the proposed method can significantly increase the fault diagnosis accuracy under imbalanced data conditions, as depicted in the experimental section (Section 5.3). Other important advantages of the proposed method are that it is more stable, robust and easier to train (e.g. reaching equilibrium between generator and discriminator) than traditional GAN-based and ACGAN-based methods. However, it was noted that for some fault types, for example for the fault data is collected within 2 s or less, the performance of the proposed method does not show significantly better performance than the compared methods.

While the proposed Multi-Scale ACGAN effectively alleviates the data imbalance problem by generating synthetic samples, the generated data still have several limitations:

- **Distribution Gap:** The synthetic data may not perfectly match the distribution of real-world fault signals, especially in terms of transient features or edge cases. This mismatch could reduce generalization in real applications.
- **Limited Interpretability:** Since the generated signals are not derived from actual physical measurements, they may lack physical meaning, which could hinder their acceptance by engineers.
- **Mode Collapse Risk:** The generator may suffer from mode collapse, where it produces limited variations of samples, thus failing to capture the full diversity of fault types.
- **Dependence on Existing Data:** The quality of generated data is fundamentally limited by the quality and representativeness of the original (real) samples. If the original samples are biased or insufficient, the synthetic data will reflect those shortcomings.

These limitations suggest that while synthetic data is useful for improving classification performance under data imbalance, further research is needed to enhance their reliability and interpretability in real-world industrial scenarios.

6. Conclusions

6.1 Summary and Key Findings

This research proposed and developed a framework of the iSEMA system, addressing critical challenges in WT monitoring, fault detection, maintenance scheduling, and fault diagnosis. By integrating DT technology with advanced DL models, the iSEMA system demonstrated its potential as a robust framework for real-time evaluation of WT states.

A significant contribution of this work is the introduction of the sub-healthy state concept, which represents an intermediate state between healthy and faulty conditions. This concept enables early detection of performance degradation before faults fully manifest, allowing for proactive maintenance and reduced operational costs. The proposed system effectively identifies sub-healthy states and provides actionable insights to operators, optimizing maintenance strategies and minimizing downtime.

Additionally, this research introduced a hybrid fault diagnosis framework that combines traditional ML models, RF, with advanced DL methods, 1DCNN. This approach improved fault diagnosis accuracy, especially for multi-class fault classification tasks. Furthermore, the enhanced Multi-Scale ACGAN addressed the imbalanced data problem by generating realistic synthetic fault data. Extensive experiments validated the efficacy of these methods, demonstrating consistent performance across diverse datasets, including offshore and inland WT data and benchmark datasets for bearing fault diagnosis.

Overall, this thesis established a strong foundation for integrating DT, ML and DL technologies in WT condition monitoring and fault diagnosis, contributing both to academic knowledge and practical applications in the renewable energy sector.

6.2 Addressing the Research Gaps

This thesis systematically addressed the key research gaps identified in Chapter 1:

- 1) **Lack of a unified and intelligent framework for real-time state evaluation and maintenance scheduling.**

Addressed by designing and implementing the iSEMA system, which integrates a DT framework with LSTM-based models for real-time state evaluation and predictive maintenance.

- 2) **Limited treatment and formalization of the sub-healthy state in WTs.**

Addressed by introducing and formalizing the concept of the sub-healthy state in WTs and proposing quantifiable indicators to detect and act upon such states in advance.

- 3) **Heavy dependence on manual feature extraction in traditional ML-based diagnosis.**

Addressed through a hybrid diagnostic model that leverages 1DCNN for automatic feature learning, combined with RF for improved classification.

- 4) **Diagnosis performance degradation under imbalanced data scenarios.**

Addressed by proposing a Multi-Scale ACGAN model to generate synthetic fault samples, thus improving diagnosis performance on imbalanced datasets.

6.3 Training Time and Computational Cost

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3060 Ti GPU and an AMD Ryzen 7 5700G CPU. The training time varied depending on the model and dataset used.

- The LSTM-based virtual model within the iSEMA framework required 7–12 minutes on the offshore/inland datasets.
- For the hybrid 1DCNN-RF model, training took approximately 15 minutes on the CWRU dataset and JNU dataset.
- For the Multi-Scale ACGAN, due to its generator-discriminator architecture and parallel FC paths, the training process took around 45 minutes for 5000 epochs on the TVRS dataset and CWRU dataset.

6.4 Practical Potential of the Proposed Methods

The proposed methods, including the iSEMA system, the fault diagnosis method based on the

hybrid 1DCNN-RF model and the Multi-Scale ACGAN-based fault diagnosis framework, exhibit strong potential for real-world deployment in WT maintenance and monitoring systems. The iSEMA system provides interpretable health indicators and early warnings, enabling maintenance teams to act before critical failures occur. This predictive capability can reduce unplanned downtime and lower maintenance costs, especially in offshore or remote wind farms where access is costly and infrequent.

Moreover, the hybrid 1DCNN-RF model offers a lightweight and computationally efficient solution, suitable for deployment on edge devices in industrial environments. Its balance between accuracy and interpretability makes it appealing to engineers who require both reliable diagnosis and transparent decision-making.

Finally, the Multi-Scale ACGAN method addresses a key industry challenge, namely data imbalance, by generating realistic synthetic data to improve model training without requiring additional physical experiments. This makes the method especially valuable for new wind farms where historical fault data are scarce.

In summary, these methods align well with the current trend toward intelligent, data-driven, and proactive maintenance strategies in modern wind energy systems.

6.5 Directions for Future Research

Based on the achievements of this study, several promising directions for future research can be identified.

One key area for further exploration is the enhancement of the iSEMA system to include fault diagnosis and prediction capabilities for other WT subsystems, such as yaw systems, hydraulic systems, and gearboxes. These subsystems are critical to WT operation, and their failures can lead to significant downtime and maintenance costs. Expanding the system's scope to these components will require designing specialized models and integrating diverse data types for better generalization.

Another direction involves experimenting with alternative sensor data to improve fault detection and diagnosis accuracy. Potential data sources include ultrasonic signals, temperature measurements, oil particle analysis, and acoustic emissions. Utilizing these data types can offer complementary perspectives and enable a more holistic assessment of WT health under varied operational conditions.

The application of the iSEMA system to multi-wind farm scenarios is another critical direction.

Real-world wind farms often involve diverse turbine configurations and environmental conditions, presenting challenges such as data standardization, scalability, and cross-farm integration. Addressing these challenges will further validate the system's adaptability and robustness.

Finally, efforts should be directed toward real-world deployment and validation of the iSEMA system. This involves collaboration with wind farm operators to implement the system in operational settings, gather real-time feedback, and refine the models based on practical insights. Concurrently, addressing challenges such as data privacy, security, and real-time processing will be essential for the large-scale adoption of the system.

By pursuing these directions, future research can further solidify the role of the iSEMA system in improving WT reliability, operational efficiency, and sustainability.

Reference

- [1] A. Rahman, O. Farrok, and M. M. Haque, "Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic," *Renew. Sust. Energ. Rev.*, Article vol. 161, p. 29, Jun 2022, Art no. 112279, doi: 10.1016/j.rser.2022.112279.
- [2] GWEC. "Gwec's global wind report 2024." <https://gwec.net/global-wind-report-2024-record-year-for-wind-energy-shows-momentum-but-highlights-need-for-policy-driven-action/> (accessed).
- [3] B. Pang, T. Tian, and G. J. Tang, "Fault state recognition of wind turbine gearbox based on generalized multi-scale dynamic time warping," *Struct. Health Monit.*, vol. 20, no. 6, pp. 3007-3023, Nov 2021, Art no. 1475921720978622, doi: 10.1177/1475921720978622.
- [4] M. Albadi. "Main components of a horizontal axis wind turbine." https://www.researchgate.net/figure/5-Main-components-of-a-horizontal-axis-wind-turbine_fig3_235340138 (accessed).
- [5] A. Bensalah, G. Barakat, and Y. Amara, "Electrical generators for large wind turbine: Trends and challenges," *Energies*, vol. 15, no. 18, Sep 2022, Art no. 6700, doi: 10.3390/en15186700.
- [6] H. Badihi, Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja, "A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis," *Proceedings of the IEEE*, vol. 110, pp. 754-806, 2022-05-13 2022, doi: 10.1109/JPROC.2022.3171691.
- [7] L. I. Gang, Q. I. Ying, L. I. Yinqiang, Z. Jianfu, and Z. Lihui, "Research progress on fault diagnosis and state prediction of wind turbine," *Automation of Electric Power Systems*, vol. 45, no. 4, pp. 180-191. [Online]. Available: aepsen/article/abstract/20200301002.
- [8] A. Stetco *et al.*, "Machine learning methods for wind turbine condition monitoring: A review," *Renew. Energy*, vol. 133, pp. 620-635, Apr 2019, doi: 10.1016/j.renene.2018.10.047.
- [9] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, vol. 5, no. 4, Dec 2017, Art no. 21, doi: 10.3390/machines5040021.
- [10] A. Widodo and B. S. Yang, "Wavelet support vector machine for induction machine fault diagnosis based on transient current signal," *Expert Syst. Appl.*, Article vol. 35, no. 1-2, pp. 307-316, Jul-Aug 2008, doi: 10.1016/j.eswa.2007.06.018.
- [11] Z. P. Feng, M. Liang, and F. L. Chu, "Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mechanical Systems and Signal Processing*, Review vol. 38, no. 1, pp. 165-205, Jul 2013, doi: 10.1016/j.ymssp.2013.01.017.
- [12] J. Y. Yang, Y. Y. Zhang, and Y. S. Zhu, "Intelligent fault diagnosis of rolling element bearing based on svms and fractal dimension," *Mechanical Systems and Signal Processing*, Article vol. 21, no. 5, pp. 2012-2024, Jul 2007, doi: 10.1016/j.ymssp.2006.10.005.
- [13] Y. B. Jing *et al.*, "Diesel engine valve clearance fault diagnosis based on features extraction techniques and fastica-svm," *Chin. J. Mech. Eng.*, Article vol. 30, no. 4, pp. 991-1007, Jul 2017, doi: 10.1007/s10033-017-0140-2.
- [14] R. Q. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mechanical Systems and Signal Processing*, Article vol. 21, no. 2, pp. 824-839,

Feb 2007, doi: 10.1016/j.ymssp.2006.02.009.

- [15] P. P. Kanjilal and S. Palit, "On multiple pattern extraction using singular-value decomposition," *IEEE Trans. Signal Process.*, Letter vol. 43, no. 6, pp. 1536-1540, Jun 1995, doi: 10.1109/78.388873.
- [16] Z. H. Zhu *et al.*, "Time series mining based on multilayer piecewise aggregate approximation," *Proceedings of 2016 International Conference on Audio, Language and Image Processing (Icalip)*, Proceedings Paper pp. 174-179, 2016. [Online]. Available: [≤Go to ISI>://WOS:000400617600034](https://doi.org/10.1109/ICALIP.2016.7800034).
- [17] F. Y. Cong, J. Chen, and Y. N. Pan, "Kolmogorov-smirnov test for rolling bearing performance degradation assessment and prognosis," *J. Vib. Control*, Article vol. 17, no. 9, pp. 1337-1347, Aug 2011, doi: 10.1177/1077546310384003.
- [18] B. Venkatesh and J. Anuradha, "A review of feature selection and its methods," *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3-26, 2019, doi: 10.2478/cait-2019-0001.
- [19] V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, Review vol. 34, no. 3, pp. 483-519, Mar 2013, doi: 10.1007/s10115-012-0487-8.
- [20] Z. H. Liu *et al.*, "A stacked auto-encoder based partial adversarial domain adaptation model for intelligent fault diagnosis of rotating machines," *IEEE Trans. Ind. Inform.*, Article vol. 17, no. 10, pp. 6798-6809, Oct 2021, doi: 10.1109/tii.2020.3045002.
- [21] W. Q. Zhao, Y. F. Zhang, and Y. L. Zhu, "Diagnosis for transformer faults based on combinatorial bayes network," *Proceedings of the 2009 2nd International Congress on Image and Signal Processing, Vols 1-9*, Proceedings Paper pp. 2563-2565, 2009. [Online]. Available: [≤Go to ISI>://WOS:000280804301191](https://doi.org/10.1109/ICISP.2009.5305191).
- [22] X. T. Zhu, J. B. Xiong, and Q. Liang, "Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm," *Ieee Access*, Article vol. 6, pp. 33583-33588, 2018, doi: 10.1109/access.2018.2789933.
- [23] T. Wumaier, C. Xu, H. Y. Guo, Z. J. Jin, and H. J. Zhou, "Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm," *Ieee Access*, Article vol. 9, pp. 69307-69315, 2021, doi: 10.1109/access.2021.3075547.
- [24] X. A. Yan and M. P. Jia, "A novel optimized svm classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, Article vol. 313, pp. 47-64, Nov 2018, doi: 10.1016/j.neucom.2018.05.002.
- [25] H. M. Yan, H. N. Mu, X. J. Yi, Y. Y. Yang, and G. L. Chen, "Fault diagnosis of wind turbine based on pca and gsa-svm," in *Prognostics and System Health Management Conference in Paris (PHM Paris)*, Paris, FRANCE, May 02-05 2019, NEW YORK: Ieee, in Prognostics and System Health Management Conference, 2019, pp. 13-17, doi: 10.1109/PHM-Paris.2019.00010. [Online]. Available: [≤Go to ISI>://WOS:000485048900003](https://doi.org/10.1109/PHM-Paris.2019.00010)
- [26] Y. Shao, X. F. Yuan, C. J. Zhang, Y. Song, and Q. Y. Xu, "A novel fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network and inpsosvm," *Applied Sciences-Basel*, Article vol. 10, no. 12, p. 21, Jun 2020, Art no. 4303, doi: 10.3390/app10124303.
- [27] L. J. Wang, S. F. Ji, and N. Y. Ji, "Comparison of support vector machine-based techniques for detection of bearing faults," *Shock and Vibration*, Article vol. 2018, p. 13, 2018, Art no. 8174860, doi: 10.1155/2018/8174860.

- [28] B. Kumar, O. P. Vyas, and R. Vyas, "A comprehensive review on the variants of support vector machines," *Modern Physics Letters B*, vol. 33, no. 25, Sep 2019, Art no. 1950303, doi: 10.1142/s0217984919503032.
- [29] T. D. Sun *et al.*, "Fault diagnosis methods based on machine learning and its applications for wind turbines: A review," *Ieee Access*, Review vol. 9, pp. 147481-147511, 2021, doi: 10.1109/access.2021.3124025.
- [30] B. Talekar and S. Agrawal, "A detailed review on decision tree and random forest," *Biosci. Biotechnol. Res. Commun.*, Review vol. 13, no. 14, pp. 245-248, 2020, doi: 10.21786/bbrc/13.14/57.
- [31] L. Canete-Sifuentes, R. Monroy, and M. A. Medina-Perez, "A review and experimental comparison of multivariate decision trees," *Ieee Access*, Review vol. 9, pp. 110451-110479, 2021, doi: 10.1109/access.2021.3102239.
- [32] O. Loyola-González, "Black-box vs. White-box: Understanding their advantages and weaknesses from a practical point of view," *IEEE ACCESS*, vol. 7, pp. 154096-154113, 2019-01-01 2019, doi: 10.1109/ACCESS.2019.2949286.
- [33] V. T. Tran, B. S. Yang, M. S. Oh, and A. C. C. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Syst. Appl.*, Article vol. 36, no. 2, pp. 1840-1849, Mar 2009, doi: 10.1016/j.eswa.2007.12.010.
- [34] L. J. Wan *et al.*, "An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm," *Ieee Access*, Article vol. 9, pp. 37866-37882, 2021, doi: 10.1109/access.2021.3063929.
- [35] X. Wang *et al.*, "Deep convolutional tree-inspired network: A decision-tree-structured neural network for hierarchical fault diagnosis of bearings," *Front. Mech. Eng.*, Article; Early Access p. 15, doi: 10.1007/s11465-021-0650-6.
- [36] H. A. Abu Alfeilat *et al.*, "Effects of distance measure choice on k-nearest neighbor classifier performance: A review," *Big Data*, Review vol. 7, no. 4, pp. 221-248, Dec 2019, doi: 10.1089/big.2018.0175.
- [37] P. Kumar and A. S. Hati, "Review on machine learning algorithm based fault detection in induction motors," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1929-1940, May 2021, doi: 10.1007/s11831-020-09446-w.
- [38] B. Song, S. Tan, H. B. Shi, and B. Zhao, "Fault detection and diagnosis via standardized k nearest neighbor for multimode process," *J. Taiwan Inst. Chem. Eng.*, Article vol. 106, pp. 1-8, Jan 2020, doi: 10.1016/j.jtice.2019.09.017.
- [39] C. Zhang, X. W. Gao, Y. Li, and L. W. Feng, "Fault detection strategy based on weighted distance of k nearest neighbors for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, Article vol. 32, no. 1, pp. 75-81, Feb 2019, doi: 10.1109/tsm.2018.2857818.
- [40] J. T. Lu, W. W. Qian, S. M. Li, and R. Q. Cui, "Enhanced k-nearest neighbor for intelligent fault diagnosis of rotating machinery," *Applied Sciences-Basel*, Article vol. 11, no. 3, p. 14, Feb 2021, Art no. 919, doi: 10.3390/app11030919.
- [41] Okfalisa, Mustakim, I. Gazalba, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," *2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (Icitisee): Opportunities and Challenges on Big Data Future Innovation*, Proceedings Paper pp. 294-298, 2017. [Online]. Available: [<Go to](#)

[ISI>://WOS:000426942900056](#).

- [42] A. Amiruddin, H. Zabiri, S. A. A. Taqvi, and L. D. Tufa, "Neural network applications in fault diagnosis and detection: An overview of implementations in engineering-related systems," *Neural Comput. Appl.*, Review vol. 32, no. 2, pp. 447-472, Jan 2020, doi: 10.1007/s00521-018-3911-5.
- [43] R. Liu, "Research on fault diagnosis based on artificial neural network," *Communications, Signal Processing, and Systems, Csps 2018, Vol Iii: Systems*, Proceedings Paper vol. 517, pp. 73-79, 2020, doi: 10.1007/978-981-13-6508-9_10.
- [44] K. Li, L. Su, J. J. Wu, H. Q. Wang, and P. Chen, "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine," *Applied Sciences-Basel*, Article vol. 7, no. 10, p. 20, Oct 2017, Art no. 1004, doi: 10.3390/app7101004.
- [45] M. L. Fadda and A. Moussaoui, "Hybrid som-pca method for modeling bearing faults detection and diagnosis," *J. Braz. Soc. Mech. Sci. Eng.*, Article vol. 40, no. 5, p. 8, May 2018, Art no. 268, doi: 10.1007/s40430-018-1184-7.
- [46] X. J. Wan, L. C. Liu, Z. B. Xu, and Z. G. Xu, "Gearbox fault diagnosis based on selective integrated soft competitive learning fuzzy adaptive resonance theory," *J. Comput. Inf. Sci. Eng.*, Article vol. 19, no. 1, p. 13, Mar 2019, Art no. 011008, doi: 10.1115/1.4041776.
- [47] S. R. Saufi, Z. A. Bin Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *Ieee Access*, Review vol. 7, pp. 122644-122662, 2019, doi: 10.1109/access.2019.2938227.
- [48] R. Zhao *et al.*, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, Article vol. 115, pp. 213-237, Jan 2019, doi: 10.1016/j.ymssp.2018.05.050.
- [49] J. Y. Jiao, M. Zhao, J. Lin, and K. X. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, Review vol. 417, pp. 36-63, Dec 2020, doi: 10.1016/j.neucom.2020.07.088.
- [50] K. Zhang *et al.*, "Residual networks of residual networks: Multilevel residual networks," *IEEE Trans. Circuits Syst. Video Technol.*, Article vol. 28, no. 6, pp. 1303-1314, Jun 2018, doi: 10.1109/tcsvt.2017.2654543.
- [51] P. Xiong, B. P. Tang, L. Deng, M. H. Zhao, and X. X. Yu, "Multi-block domain adaptation with central moment discrepancy for fault diagnosis," *Measurement*, Article vol. 169, p. 10, Feb 2021, Art no. 108516, doi: 10.1016/j.measurement.2020.108516.
- [52] P. Singh and G. Pradhan, "A new ecg denoising framework using generative adversarial network," *IEEE-ACM Trans. Comput. Biol. Bioinform.*, Article vol. 18, no. 2, pp. 759-764, Mar-Apr 2021, doi: 10.1109/tcbb.2020.2976981.
- [53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning*, 2015, vol. 37, 2015, pp. 448-456.
- [54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, Article vol. 15, pp. 1929-1958, Jun 2014. [Online]. Available: [<Go to ISI>://WOS:000344638300002](#).
- [55] W. F. Gong *et al.*, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved cnn-svm and multichannel data fusion," *Sensors*, Article vol.

- 19, no. 7, p. 37, Apr 2019, Art no. 1693, doi: 10.3390/s19071693.
- [56] F. Wang, H. G. Jiang, H. D. Shao, W. J. Duan, and S. P. Wu, "An adaptive deep convolutional neural network for rolling bearing fault diagnosis," *Meas. Sci. Technol.*, Article vol. 28, no. 9, p. 17, Sep 2017, Art no. 095005, doi: 10.1088/1361-6501/aa6e22.
 - [57] W. T. Zhang, D. Yang, H. C. Wang, X. F. Huang, and M. Gidlund, "Carnet: A dual correlation method for health perception of rotating machinery," *IEEE Sens. J.*, Article vol. 19, no. 16, pp. 7095-7106, Aug 2019, doi: 10.1109/jsen.2019.2912934.
 - [58] H. Q. Wang, S. Li, L. Y. Song, L. L. Cui, and P. X. Wang, "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," *IEEE Trans. Instrum. Meas.*, Article vol. 69, no. 6, pp. 2648-2657, Jun 2020, doi: 10.1109/tim.2019.2928346.
 - [59] V. Pandhare, J. Singh, and J. Lee, "Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features," in *Prognostics and System Health Management Conference in Paris (PHM Paris)*, 2019, 2019, pp. 320-326, doi: 10.1109/PHM-Paris.2019.00061.
 - [60] Y. G. Lei *et al.*, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, Apr 2020, Art no. 106587, doi: 10.1016/j.ymssp.2019.106587.
 - [61] W. W. Qian, S. M. Li, J. R. Wang, Z. H. An, and X. X. Jiang, "An intelligent fault diagnosis framework for raw vibration signals: Adaptive overlapping convolutional neural network," *Meas. Sci. Technol.*, Article vol. 29, no. 9, p. 13, Sep 2018, Art no. 095009, doi: 10.1088/1361-6501/aad101.
 - [62] C. C. Chen, Z. Liu, G. S. Yang, C. C. Wu, and Q. B. Ye, "An improved fault diagnosis using 1d-convolutional neural network model," *Electronics*, Article vol. 10, no. 1, p. 19, Jan 2021, Art no. 59, doi: 10.3390/electronics10010059.
 - [63] S. Shen *et al.*, "A physics-informed deep learning approach for bearing fault detection," *Eng. Appl. Artif. Intell.*, Article vol. 103, p. 15, Aug 2021, Art no. 104295, doi: 10.1016/j.engappai.2021.104295.
 - [64] R. Y. Huang, Y. X. Liao, S. H. Zhang, and W. H. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *Ieee Access*, Article vol. 7, pp. 1848-1858, 2019, doi: 10.1109/access.2018.2886343.
 - [65] Q. P. Chen, Q. S. Xie, Q. N. Yuan, H. S. Huang, and Y. T. Li, "Research on a real-time monitoring method for the wear state of a tool based on a convolutional bidirectional lstm model," *Symmetry-Basel*, Article vol. 11, no. 10, p. 18, Oct 2019, Art no. 1233, doi: 10.3390/sym11101233.
 - [66] X. Y. Li, J. L. Li, Y. Z. Qu, and D. He, "Gear pitting fault diagnosis using integrated cnn and gru network with both vibration and acoustic emission signals," *Applied Sciences-Basel*, Article vol. 9, no. 4, p. 15, Feb 2019, Art no. 768, doi: 10.3390/app9040768.
 - [67] I. H. Kao, W. J. Wang, Y. H. Lai, and J. W. Perng, "Analysis of permanent magnet synchronous motor fault diagnosis based on learning," *IEEE Trans. Instrum. Meas.*, Article vol. 68, no. 2, pp. 310-324, Feb 2019, doi: 10.1109/tim.2018.2847800.
 - [68] T. Khan, P. Alekhya, J. Seshadrinath, and Ieee, "Incipient inter-turn fault diagnosis in induction motors using cnn and lstm based methods," *2018 Ieee Industry Applications Society Annual Meeting (Ias)*, Proceedings Paper p. 6, 2018. [Online]. Available: [<Go to](#)

ISI>://WOS:000719813000196.

- [69] S. Kiranyaz *et al.*, "1d convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, Article vol. 151, p. 21, Apr 2021, Art no. 107398, doi: 10.1016/j.ymssp.2020.107398.
- [70] J. J. Zhu *et al.*, "Application of recurrent neural network to mechanical fault diagnosis: A review," *J. Mech. Sci. Technol.*, Review; Early Access p. 16, doi: 10.1007/s12206-022-0102-1.
- [71] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, Article vol. 9, no. 8, pp. 1735-1780, Nov 1997, doi: 10.1162/neco.1997.9.8.1735.
- [72] Z. H. Liu *et al.*, "A regularized lstm method for predicting remaining useful life of rolling bearings," *Int. J. Autom. Comput.*, Article vol. 18, no. 4, pp. 581-593, Aug 2021, doi: 10.1007/s11633-020-1276-6.
- [73] H. Ouyang, J. S. Zeng, Y. F. Li, and S. H. Luo, "Fault detection and identification of blast furnace ironmaking process using the gated recurrent unit network," *Processes*, Article vol. 8, no. 4, p. 15, Apr 2020, Art no. 391, doi: 10.3390/pr8040391.
- [74] D. Cabrera *et al.*, "Bayesian approach and time series dimensionality reduction to lstm-based model-building for fault diagnosis of a reciprocating compressor," *Neurocomputing*, Article vol. 380, pp. 51-66, Mar 2020, doi: 10.1016/j.neucom.2019.11.006.
- [75] X. D. Miao, S. M. Li, Y. Q. Zhu, and Z. H. An, "A novel real-time fault diagnosis method for planetary gearbox using transferable hidden layer," *IEEE Sens. J.*, Article vol. 20, no. 15, pp. 8403-8412, Aug. Jan 2020, doi: 10.1109/jsen.2020.2965988.
- [76] Z. Y. Xue, K. S. Xiahou, M. S. Li, T. Y. Ji, and Q. H. Wu, "Diagnosis of multiple open-circuit switch faults based on long short-term memory network for dfig-based wind turbine systems," *IEEE J. Emerg. Sel. Top. Power Electron.*, Article vol. 8, no. 3, pp. 2600-2610, Sept 2020, doi: 10.1109/jestpe.2019.2908981.
- [77] M. Y. Qiao, S. H. Yan, X. X. Tang, and C. K. Xu, "Deep convolutional and lstm recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads," *Ieee Access*, Article vol. 8, pp. 66257-66269, 2020, doi: 10.1109/access.2020.2985617.
- [78] T. Lei, Y. Zhang, and Y. Artzi, "Training rnns as fast as cnns," *ArXiv*, vol. abs/1709.02755, 2017.
- [79] L. X. Cao, Z. Qian, H. Zareipour, Z. K. Huang, and F. H. Zhang, "Fault diagnosis of wind turbine gearbox based on deep bi-directional long short-term memory under time-varying non-stationary operating conditions," *Ieee Access*, Article vol. 7, pp. 155219-155228, 2019, doi: 10.1109/access.2019.2947501.
- [80] P. Park, P. Di Marco, H. Shin, and J. Bang, "Fault detection and diagnosis using combined autoencoder and long short-term memory network," *Sensors*, Article vol. 19, no. 21, p. 17, Nov 2019, Art no. 4612, doi: 10.3390/s19214612.
- [81] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, Article vol. 72-73, pp. 303-315, May 2016, doi: 10.1016/j.ymssp.2015.10.025.
- [82] W. N. Lu, X. Q. Wang, C. C. Yang, T. Zhang, and Ieee, "A novel feature extraction method using deep neural network for rolling bearing fault diagnosis," in *27th Chinese Control and Decision Conference (CCDC)*, 2015, 2015, pp. 2427-2431.

- [83] Y. Y. Long, W. N. Zhou, and Y. Luo, "A fault diagnosis method based on one-dimensional data enhancement and convolutional neural network," *Measurement*, Article vol. 180, p. 10, Aug 2021, Art no. 109532, doi: 10.1016/j.measurement.2021.109532.
- [84] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, Article vol. 11, pp. 3371-3408, Dec 2010. [Online]. Available: [Go to ISI>://WOS:000286637200004](https://www.isi.edu/WOS/000286637200004).
- [85] B. P. Duong and J. M. Kim, "Fault diagnosis of multiple combined defects in bearings using a stacked denoising autoencoder," *Advances in Computer Communication and Computational Sciences, Vol I*, Proceedings Paper vol. 759, pp. 83-93, 2019, doi: 10.1007/978-981-13-0341-8_8.
- [86] F. Xu and P. W. Tse, "Automatic roller bearings fault diagnosis using dsae in deep learning and cfs algorithm," *Soft Comput.*, Article vol. 23, no. 13, pp. 5117-5128, Jul 2019, doi: 10.1007/s00500-018-3178-x.
- [87] M. A. Ranzato, C. Poultney, S. Chopra, and Y. Cun, "Efficient learning of sparse representations with an energy-based model," *Advances in neural information processing systems*, vol. 19, 2006.
- [88] Y. M. Qi *et al.*, "An automatic feature learning and fault diagnosis method based on stacked sparse autoencoder," *Advanced Manufacturing and Automation Vii*, Proceedings Paper vol. 451, pp. 367-375, 2018, doi: 10.1007/978-981-10-5768-7_39.
- [89] L. W. Zhang *et al.*, "A review on deep learning applications in prognostics and health management," *Ieee Access*, Review vol. 7, pp. 162415-162438, 2019, doi: 10.1109/access.2019.2950985.
- [90] C. Q. Shen, Y. M. Qi, J. Wang, G. G. Cai, and Z. K. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive," *Eng. Appl. Artif. Intell.*, Article vol. 76, pp. 170-184, Nov 2018, doi: 10.1016/j.engappai.2018.09.010.
- [91] S. Mushtaq, M. M. M. Islam, and M. Sohaib, "Deep learning aided data-driven fault diagnosis of rotatory machine: A comprehensive review," *Energies*, Review vol. 14, no. 16, p. 24, Aug 2021, Art no. 5150, doi: 10.3390/en14165150.
- [92] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, pp. 139-144, 2020-11-01 2020, doi: 10.1145/3422622.
- [93] Y. Chuang, H. Hsu, K. Lin, R. Chang, and H. Lee, "Metaex-gan: Meta exploration to improve natural language generation via generative adversarial networks," *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 31, pp. 3968-3980, 2023-01-01 2023, doi: 10.1109/TASLP.2023.3317571.
- [94] Y. Cao *et al.*, "Recent advances of generative adversarial networks in computer vision," *IEEE ACCESS*, vol. 7, pp. 14985-15006, 2019-01-01 2019, doi: 10.1109/ACCESS.2018.2886814.
- [95] S. Altun and M. Talu, "A new approach for pap-smear image generation with generative adversarial networks," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 37, no. 3, pp. 1401-1410, 2022-01-01 2022, doi: 10.17341/gazimmfd.939092.
- [96] Y. Hono *et al.*, "Singing voice synthesis based on generative adversarial networks," in *2019 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2019-01-01 2019, pp. 6955-6959, doi: 10.1109/icassp.2019.8683154.
- [97] A. Karjalainen, R. Mitchell, J. Vazquez, and IEEE, "Training and validation of automatic

- target recognition systems using generative adversarial networks," in *2019 Sensor Signal Processing for Defence Conference (SSPD)*, 2019-01-01 2019, doi: 10.1109/sspd.2019.8751666.
- [98] S. Nam, S. Jeon, and J. Moon, "A new password cracking model with generative adversarial networks," in *Information Security Applications, WISA 2019*, 2020-01-01 2020, vol. 11897, pp. 247-258, doi: 10.1007/978-3-030-39303-8_19.
- [99] J. McGinn, C. Messenger, M. Williams, and I. Heng, "Generalised gravitational wave burst generation with generative adversarial networks," *Classical and Quantum Gravity*, vol. 38, no. 15, 2021-08-05 2021, Art no. 155005, doi: 10.1088/1361-6382/ac09cc.
- [100] R. Narikawa *et al.*, "Generative adversarial networks-based synthetic microstructures for data-driven materials design," *Advanced Theory and Simulations*, vol. 5, no. 5, 2022-02-20 2022, Art no. 2100470, doi: 10.1002/adts.202100470.
- [101] J. Cheng *et al.*, "Generative adversarial networks: A literature review," *KSII Transactions on Internet and Information Systems*, vol. 14, pp. 4625-4647, 2020-12-31 2020, doi: 10.3837/tiis.2020.12.001.
- [102] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213-222, 2018-10-08 2018, doi: 10.1016/j.neucom.2018.05.024.
- [103] D. Ruan, X. Song, C. Gühmann, and J. Yan, "Collaborative optimization of cnn and gan for bearing fault diagnosis under unbalanced datasets," *Lubricants*, vol. 9, 2021-10-01 2021, Art no. 105, doi: 10.3390/lubricants9100105.
- [104] X. Xu, X. Chen, and Y. Zhao, "Data imbalance bearing fault diagnosis based on fusion attention mechanism and global feature cross gan network," *Meas. Sci. Technol.*, vol. 35, 2024-10-01 2024, Art no. 106136, doi: 10.1088/1361-6501/ad64f5.
- [105] M. Pham, J. Kim, and C. Kim, "Rolling bearing fault diagnosis based on improved gan and 2-d representation of acoustic emission signals," *IEEE ACCESS*, vol. 10, pp. 78056-78069, 2022-01-01 2022, doi: 10.1109/ACCESS.2022.3193244.
- [106] T. Kim *et al.*, "Spectrum-guided gan with density-directionality sampling: Diverse high-fidelity signal generation for fault diagnosis of rotating machinery," *Advanced Engineering Informatics*, vol. 62, 2024-10-21 2024, Art no. 102821, doi: 10.1016/j.aei.2024.102821.
- [107] J. Seon *et al.*, "Least information spectral gan with time-series data augmentation for industrial iot," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024-06-04 2024, doi: 10.1109/TETCI.2024.3406719.
- [108] S. Afrasiabi *et al.*, "Wind turbine fault diagnosis with generative-temporal convolutional neural network," in *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 2019-01-01 2019, doi: 10.1109/eeeic.2019.8783233.
- [109] S. Lee, H. Jeong, and J. Kwon, "Transformer-based gan with multi-stft for rotating machinery vibration data analysis," *Electronics*, vol. 13, no. 21, 2024-11-01 2024, Art no. 4253, doi: 10.3390/electronics13214253.
- [110] M. Hasan, S. Jan, and I. Koo, "Wasserstein gan-based digital twin-inspired model for early drift fault detection in wireless sensor networks," *IEEE Sens. J.*, vol. 23, no. 12, pp. 13327-13339, 2023-06-15 2023, doi: 10.1109/JSEN.2023.3272908.

- [111] M. Rathore and S. Harsha, "Framework for imbalanced fault diagnosis of rolling bearing using autoencoding generative adversarial learning," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 45, no. 1, 2023-01-01 2023, Art no. 39, doi: 10.1007/s40430-022-03955-4.
- [112] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International Conference on Machine Learning*, 2017-01-01 2017, vol. 70.
- [113] D. Ruan, X. Chen, C. Gühmann, and J. Yan, "Improvement of generative adversarial network and its application in bearing fault diagnosis: A review," *Lubricants*, vol. 11, 2023-02-01 2023, Art no. 74, doi: 10.3390/lubricants11020074.
- [114] P. Peng *et al.*, "Imbalanced process fault diagnosis using enhanced auxiliary classifier gan," in *2020 Chinese Automation Congress (CAC 2020)*, 2020-01-01 2020, pp. 313-316, doi: 10.1109/CAC51589.2020.9327104.
- [115] Z. Fu, Z. Liu, S. Ping, W. Li, and J. Liu, "Tra-acgan: A motor bearing fault diagnosis model based on an auxiliary classifier generative adversarial network and transformer network," *ISA Transactions*, vol. 149, pp. 381-393, 2024-06-01 2024, doi: 10.1016/j.isatra.2024.03.033.
- [116] J. Chen, C. Lin, J. Cui, and H. Ge, "An fault diagnostic method based on drn-acgan for data imbalance," in *2022 Prognostics and Health Management Conference, PHM-LONDON 2022*, 2022-01-01 2022, pp. 97-102, doi: 10.1109/PHM2022-London52454.2022.00025.
- [117] W. Li, X. Zhong, H. Shao, B. Cai, and X. Yang, "Multi-mode data augmentation and fault diagnosis of rotating machinery using modified acgan designed with new framework," *Advanced Engineering Informatics*, vol. 52, 2022-02-16 2022, Art no. 101552, doi: 10.1016/j.aei.2022.101552.
- [118] C. Wu and Z. Zeng, "A fault diagnosis method based on auxiliary classifier generative adversarial network for rolling bearing," *Plos One*, vol. 16, 2021-03-01 2021, Art no. e0246905, doi: 10.1371/journal.pone.0246905.
- [119] S. Dixit, N. Verma, and A. Ghosh, "Intelligent fault diagnosis of rotary machines: Conditional auxiliary classifier gan coupled with meta learning using limited data," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021-01-01 2021, doi: 10.1109/TIM.2021.3082264.
- [120] T. Kari *et al.*, "Fault diagnosis of power transformer based on improved acgan under imbalanced data," *CMC-Computers, Materials & Continua*, vol. 75, no. 2, pp. 4573-4592, 2023-01-01 2023, doi: 10.32604/cmc.2023.037954.
- [121] H. Zhong *et al.*, "Fine-tuning transfer learning based on dcgan integrated with self-attention and spectral normalization for bearing fault diagnosis," *Measurement*, Article vol. 210, 2023-02-03 2023, Art no. 112421, doi: 10.1016/j.measurement.2022.112421.
- [122] Y. He, H. Tang, Y. Ren, and A. Kumar, "A semi-supervised fault diagnosis method for axial piston pump bearings based on dcgan," *Meas. Sci. Technol.*, Article vol. 32, no. 12, 2021-12-01 2021, Art no. 125104, doi: 10.1088/1361-6501/ac1fbc.
- [123] L. Guo *et al.*, "Unsupervised feature-preserving cyclegan for fault diagnosis of rolling bearings using unbalanced infrared thermal imaging sample," *IEEE ACCESS*, Article vol. 12, pp. 28449-28461, 2024-01-01 2024, doi: 10.1109/ACCESS.2024.3365551.
- [124] J. Zhao and W. Huang, "Transfer learning method for rolling bearing fault diagnosis under different working conditions based on cyclegan," *Meas. Sci. Technol.*, Article vol. 33, no. 2, 2022-02-01 2022, Art no. 025003, doi: 10.1088/1361-6501/ac3942.
- [125] J. Chen, Z. Yan, C. Lin, B. Yao, and H. Ge, "Aero-engine high speed bearing fault diagnosis for data imbalance: A sample enhanced diagnostic method based on pre-training wgan-gp,"

- Measurement*, Article vol. 213, 2023-03-20 2023, Art no. 112709, doi: 10.1016/j.measurement.2023.112709.
- [126] H. Fan *et al.*, "Intelligent data expansion approach of vibration gray texture images of rolling bearing based on improved wgan-gp," *Adv. Mech. Eng.*, Article vol. 14, no. 3, 2022-03-01 2022, Art no. 16878132221086132, doi: 10.1177/16878132221086132.
 - [127] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," *White paper*, vol. 1, no. 2014, pp. 1-7, 2014.
 - [128] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the digital twin: A systematic literature review," *CIRP J. Manuf. Sci. Technol.*, Review vol. 29, pp. 36-52, May 2020, doi: 10.1016/j.cirpj.2020.02.002.
 - [129] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital twin in manufacturing: A categorical literature review and classification," in *IFAC-PapersOnLine*, 2018-01-01 2018, vol. 51, pp. 1016-1022, doi: 10.1016/j.ifacol.2018.08.474.
 - [130] F. Tao *et al.*, "Five-dimension digital twin model and its ten applications," *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS*, Article vol. 25, no. 1, pp. 1-18, 2019, doi: 10.13196/j.cims.2019.01.001.
 - [131] F. Tao, B. Xiao, Q. L. Qi, J. F. Cheng, and P. Ji, "Digital twin modeling," *J. Manuf. Syst.*, Review vol. 64, pp. 372-389, Jul 2022, doi: 10.1016/j.jmsy.2022.06.015.
 - [132] B. Chen, K. Liu, T. Yu, and R. Li, "Enhancing reliability in floating offshore wind turbines through digital twin technology: A comprehensive review," *Energies*, vol. 17, 2024-04-01 2024, Art no. 1964, doi: 10.3390/en17081964.
 - [133] C. Yang *et al.*, "Comprehensive analysis and evaluation of the operation and maintenance of offshore wind power systems: A survey," *Energies*, vol. 16, 2023-07-01 2023, Art no. 5562, doi: 10.3390/en16145562.
 - [134] M. Dinh *et al.*, "Implementation of digital twin-assisted condition monitoring and fault diagnosis for wind turbines," in *2023 12th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2023-01-01 2023, pp. 146-150, doi: 10.1109/ICRERA59003.2023.10269370.
 - [135] H. Liu, W. Sun, S. Bao, L. Xiao, and L. Jiang, "Research on key technology of wind turbine drive train fault diagnosis system based on digital twin," *Applied Sciences-Basel*, vol. 14, 2024-07-01 2024, Art no. 5991, doi: 10.3390/app14145991.
 - [136] I. Abdullahi, S. Longo, and M. Samie, "Towards a distributed digital twin framework for predictive maintenance in industrial internet of things (iiot)," *Sensors*, vol. 24, 2024-04-01 2024, Art no. 2663, doi: 10.3390/s24082663.
 - [137] M. Wang *et al.*, "Recent progress on reliability analysis of offshore wind turbine support structures considering digital twin solutions," *Ocean Engineering*, vol. 232, 2021-05-21 2021, Art no. 109168, doi: 10.1016/j.oceaneng.2021.109168.
 - [138] R. Pacheco-Blazquez, J. Garcia-Espinosa, D. Di Capua, and A. Sanchez, "A digital twin for assessing the remaining useful life of offshore wind turbine structures," *Journal of Marine Science and Engineering*, vol. 12, 2024-04-01 2024, Art no. 573, doi: 10.3390/jmse12040573.
 - [139] S. Lee *et al.*, "Recent progress and performance analysis on durability evaluation and remaining useful life prediction technology development for the life extension of wind turbines in korea," in *2023 12th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2023-01-01 2023, pp. 339-343, doi:

10.1109/ICRERA59003.2023.10269432.

- [140] M. Ma, Y. Hou, and Y. Li, "A multi-scale feature fusion network-based fault diagnosis method for wind turbine bearings," *Wind Engineering*, vol. 47, pp. 3-15, 2022-07-23 2023, doi: 10.1177/0309524X221114621.
- [141] F. Zhou, Z. Zhang, D. Chen, and IEEE, "Bearing fault diagnosis based on dnn using multi-scale feature fusion," in *2020 35TH Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2020-01-01 2020, pp. 150-155, doi: 10.1109/YAC51587.2020.9337689.
- [142] Y. Xie *et al.*, "Multi-scale deep neural network for fault diagnosis method of rotating machinery," *Ferroelectrics*, vol. 602, pp. 215-230, 2023-01-02 2023, doi: 10.1080/00150193.2022.2079456.
- [143] L. Deng, Y. Zhang, and Z. Shi, "An improved fault diagnosis method of rolling bearings based on multi-scale attention cnn," *Journal of Failure Analysis and Prevention*, vol. 24, pp. 1814-1827, 2024-06-20 2024, doi: 10.1007/s11668-024-01957-z.
- [144] Y. Wang, J. Liang, X. Gu, D. Ling, and H. Yu, "Multi-scale attention mechanism residual neural network for fault diagnosis of rolling bearings," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 236, pp. 10615-10629, 2022-05-30 2022, Art no. 09544062221104598, doi: 10.1177/09544062221104598.
- [145] N. Rodriguez, P. Alvarez, L. Barba, and G. Cabrera-Guerrero, "Combining multi-scale wavelet entropy and kernelized classification for bearing multi-fault diagnosis," *Entropy*, vol. 21, no. 2, 2019-02-01 2019, Art no. 152, doi: 10.3390/e21020152.
- [146] S. Kim *et al.*, "Hybrid data-scaling method for fault classification of compressors," *Measurement*, vol. 201, 2022-08-12 2022, Art no. 111619, doi: 10.1016/j.measurement.2022.111619.
- [147] C. Semeraro *et al.*, "Digital twin application in energy storage: Trends and challenges," *J. Energy Storage*, Review vol. 58, p. 15, Feb 2023, Art no. 106347, doi: 10.1016/j.est.2022.106347.
- [148] B. He and K. J. Bai, "Digital twin-based sustainable intelligent manufacturing: A review," *Adv. Manuf.*, Review vol. 9, no. 1, pp. 1-21, Mar 2021, doi: 10.1007/s40436-020-00302-5.
- [149] E. Shahat, C. T. Hyun, and C. Yeom, "City digital twin potentials: A review and research agenda," *Sustainability*, Review vol. 13, no. 6, p. 20, Mar 2021, Art no. 3386, doi: 10.3390/su13063386.
- [150] Y. Wang *et al.*, "Fault diagnosis of wind turbine planetary gear based on a digital twin," *Applied Sciences-Basel*, Article vol. 13, no. 8, p. 19, Apr 2023, Art no. 4776, doi: 10.3390/app13084776.
- [151] C. Kim *et al.*, "Design, implementation, and evaluation of an output prediction model of the 10 mw floating offshore wind turbine for a digital twin," *Energies*, Article vol. 15, no. 17, p. 16, Sep 2022, Art no. 6329, doi: 10.3390/en15176329.
- [152] X. Y. Xie *et al.*, "Fault diagnosis method for bearing based on digital twin," *Math. Probl. Eng.*, Article vol. 2022, p. 15, Nov 2022, Art no. 2982746, doi: 10.1155/2022/2982746.
- [153] Y. Guo, L. Z. Guo, S. A. Billings, and H. L. Wei, "Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems," *Neurocomputing*, Article vol. 173, pp. 715-723, Jan 2016, doi: 10.1016/j.neucom.2015.08.022.

- [154] Y. Li, H. L. Wei, S. A. Billings, and X. F. Liao, "Time-varying linear and nonlinear parametric model for granger causality analysis (vol 85, 041906, 2012)," *Phys. Rev. E*, Correction vol. 85, no. 4, p. 1, Apr 2012, Art no. 049908, doi: 10.1103/PhysRevE.85.049908.
- [155] H. L. Wei, S. A. Billings, and M. A. Balikhin, "Wavelet based non-parametric narx models for nonlinear input-output system identification," *Int. J. Syst. Sci.*, Article vol. 37, no. 15, pp. 1089-1096, Dec 2006, doi: 10.1080/00207720600903011.
- [156] R. Pandit, D. Infield, and M. Santos, "Accounting for environmental conditions in data-driven wind turbine power models," *IEEE Transactions on Sustainable Energy*, vol. 14, pp. 168-177, 2023-01-01 2023, doi: 10.1109/TSTE.2022.3204453.
- [157] Z. Lin, D. Cevasco, and M. Collu, "A methodology to develop reduced-order models to support the operation and maintenance of offshore wind turbines," *Applied Energy*, vol. 259, 2020-02-01 2020, Art no. 114228, doi: 10.1016/j.apenergy.2019.114228.
- [158] Z. H. Liu, H. L. Wei, X. H. Li, K. Liu, and Q. C. Zhong, "Global identification of electrical and mechanical parameters in pmsm drive based on dynamic self-learning pso," *IEEE Trans. Power Electron.*, Article vol. 33, no. 12, pp. 10858-10871, Dec 2018, doi: 10.1109/tpel.2018.2801331.
- [159] Z. H. Liu, H. L. Wei, Q. C. Zhong, K. Liu, and X. H. Li, "Gpu implementation of dpso-re algorithm for parameters identification of surface pmsm considering vsr nonlinearity," *IEEE J. Emerg. Sel. Top. Power Electron.*, Article vol. 5, no. 3, pp. 1334-1345, Sep 2017, doi: 10.1109/jestpe.2017.2690688.
- [160] A. J. Hu, L. Xiang, and L. J. Zhu, "An engineering condition indicator for condition monitoring of wind turbine bearings," *Wind Energy*, vol. 23, no. 2, pp. 207-219, Feb 2020, doi: 10.1002/we.2423.
- [161] G. Van Houdt, C. Mosquera, and G. Napoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, Review vol. 53, no. 8, pp. 5929-5955, Dec 2020, doi: 10.1007/s10462-020-09838-1.
- [162] P. Lei, F. Ma, C. Zhu, and T. Li, "Lstm short-term wind power prediction method based on data preprocessing and variational modal decomposition for soft sensors," *Sensors*, vol. 24, 2024-04-01 2024, Art no. 2521, doi: 10.3390/s24082521.
- [163] M. L. Hossain, A. Abu-Siada, and S. M. Mueen, "Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review," *Energies*, Review vol. 11, no. 5, p. 14, May 2018, Art no. 1309, doi: 10.3390/en11051309.
- [164] J. Zhang, N. Jiang, H. Li, and N. Li, "Online health assessment of wind turbine based on operational condition recognition," *TRANSACTIONS OF THE INSTITUTE OF MEASUREMENT AND CONTROL*, vol. 41, no. 10, pp. 2970-2981, 2019-06-01 2019, doi: 10.1177/0142331218810070.
- [165] Y. Ding. "Inland-offshore wind farm dataset2." <https://doi.org/10.5281/zenodo.5516554> (accessed).
- [166] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *Ieee Access*, Review vol. 8, pp. 93155-93178, 2020, doi: 10.1109/access.2020.2990528.
- [167] S. Y. Yoon and J. F. MacGregor, "Fault diagnosis with multivariate statistical models part i: Using steady state fault signatures," *J. Process Control*, Article vol. 11, no. 4, pp. 387-400, Aug 2001, doi: 10.1016/s0959-1524(00)00008-1.

- [168] X. Zhang, B. Y. Zhao, and Y. Lin, "Machine learning based bearing fault diagnosis using the case western reserve university data: A review," *Ieee Access*, vol. 9, pp. 155598-155608, 2021, doi: 10.1109/access.2021.3128669.
- [169] "Case western reserve university bearing data." <https://engineering.case.edu/bearingdatacenter/download-data-file> (accessed.
- [170] Y. Seo, S. Kim, O. Kisi, V. P. Singh, and K. Parasuraman, "River stage forecasting using wavelet packet decomposition and machine learning models," *Water Resour. Manag.*, Article vol. 30, no. 11, pp. 4011-4035, Sep 2016, doi: 10.1007/s11269-016-1409-4.
- [171] W. Liu, Q. Huang, M. Wei, and Ieee, "Image quality evaluation based on svm and improved grid search algorithm," *15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (Tcset - 2020)*, Proceedings Paper pp. 842-845, 2020, doi: 10.1109/tcset49122.2020.235555.
- [172] Z. L. Liu, H. Wang, J. J. Liu, Y. Qin, and D. D. Peng, "Multitask learning based on lightweight 1dcnn for fault diagnosis of wheelset bearings," *IEEE Trans. Instrum. Meas.*, Article vol. 70, p. 11, 2021, Art no. 3501711, doi: 10.1109/tim.2020.3017900.
- [173] "Jnu." <http://mad-net.org:8765/explore.html?t=0.5831516555847212> (accessed.
- [174] W. Q. Liao *et al.*, "A review: The application of generative adversarial network for mechanical fault diagnosis," *Meas. Sci. Technol.*, Review vol. 35, no. 6, p. 16, Jun 2024, Art no. 062002, doi: 10.1088/1361-6501/ad356d.
- [175] X. Li *et al.*, "A review on convolutional neural network in rolling bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 35, 2024-07-01 2024, Art no. 072002, doi: 10.1088/1361-6501/ad356e.
- [176] X. Liu, Z. Zhang, F. Meng, and Y. Zhang, "Fault diagnosis of wind turbine bearings based on cnn and ssa-elm," *Journal of Vibration Engineering & Technologies*, vol. 11, pp. 3929-3945, 2022-11-29 2023, doi: 10.1007/s42417-022-00793-5.
- [177] M. Pandiyan and T. Babu, "Systematic review on fault diagnosis on rolling-element bearing," *Journal of Vibration Engineering & Technologies*, 2024-04-10 2024, doi: 10.1007/s42417-024-01358-4.
- [178] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in Brief*, vol. 21, pp. 1745-1749, 2018-12-01 2018, doi: 10.1016/j.dib.2018.11.019.
- [179] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization gan for unbalanced data," *Knowledge-Based Systems*, vol. 187, 2020-01-01 2020, Art no. 104837, doi: 10.1016/j.knosys.2019.07.008.
- [180] Z. Zhong *et al.*, "Imbalanced bearing fault diagnosis based on rfh-gan and psa-drsn," *IEEE ACCESS*, vol. 11, pp. 131926-131938, 2023-01-01 2023, doi: 10.1109/ACCESS.2023.3335199.